

# The PAU Survey: The quasar Ly $\alpha$ and UV luminosity functions at $2.7 < z < 5.3$

Alberto Torralba-Torregrosa<sup>1,2,\*</sup>, Pablo Renard<sup>3,\*</sup>, Daniele Spinoso<sup>3</sup>, Pablo Arnalte-Mur<sup>1,2</sup>,  
Siddhartha Gurung-López<sup>1,2</sup>, Alberto Fernández-Soto<sup>4,5</sup>, Enrique Gaztañaga<sup>6,7,8</sup>, David Navarro-Gironés<sup>7,8</sup>,  
Zheng Cai<sup>3</sup>, Jorge Carretero<sup>9,10</sup>, Francisco J. Castander<sup>7,8</sup>, Martin Eriksen<sup>11,10</sup>, Juan Garcia-Bellido<sup>12</sup>,  
Hendrik Hildebrandt<sup>13</sup>, Henk Hoekstra<sup>14</sup>, Ramon Miquel<sup>11,15</sup>, Eusebio Sanchez<sup>9</sup>, Pau Tallada-Crespí<sup>9,10</sup>,  
Juan De Vicente<sup>9</sup>, and Enrique Fernandez<sup>11</sup>

- <sup>1</sup> Observatori Astronòmic de la Universitat de València, Ed. Instituts d'Investigació, Parc Científic, C/ Catedrático José Beltrán, n2, 46980 Paterna, Valencia, Spain
- <sup>2</sup> Departament d'Astronomia i Astrofísica, Universitat de València, 46100 Burjassot, Spain
- <sup>3</sup> Department of Astronomy, Tsinghua University, Beijing 100084, China
- <sup>4</sup> Instituto de Física de Cantabria (CSIC-UC), Avda. Los Castros s/n, 39005 Santander, Spain
- <sup>5</sup> Unidad Asociada "Grupo de Astrofísica Extragaláctica y Cosmología", IFCA-CSIC/Universitat de València, València, Spain
- <sup>6</sup> Institute of Cosmology and Gravitation (University of Portsmouth), Portsmouth, UK
- <sup>7</sup> Institute of Space Sciences (ICE, CSIC), E-08193 Barcelona, Spain
- <sup>8</sup> Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
- <sup>9</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, E-28040 Madrid, Spain
- <sup>10</sup> Port d'Informació Científica (PIC), Campus UAB, C. Albareda s/n, 08193 Bellaterra, Barcelona, Spain
- <sup>11</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra, Barcelona, Spain
- <sup>12</sup> Instituto de Física Teórica (UAM/CSIC), Nicolas Cabrera 13, E-28049 Madrid, Spain
- <sup>13</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
- <sup>14</sup> Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
- <sup>15</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), 0810 Barcelona, Spain

Received 26 July 2024 / Accepted 6 September 2024

## ABSTRACT

We present the Lyman- $\alpha$  (Ly $\alpha$ ) and ultraviolet (UV) luminosity function (LF) in bins of redshift for quasars selected in the Physics of the Accelerating Universe Survey (PAUS). A sample of 915 objects was selected at  $2.7 < z < 5.3$  within an effective area of  $\sim 36 \text{ deg}^2$  observed in 40 narrow-band (NB) filters (FWHM  $\sim 120 \text{ \AA}$ ). We cover the intermediate–bright luminosity regime of the LF ( $10^{43.5} < (L_{\text{Ly}\alpha}/\text{erg s}^{-1}) < 10^{45.5}$ ;  $-29 < M_{\text{UV}} < -24$ ). The continuous wavelength coverage of the PAUS NB set allows very efficient target identification and precise redshift measurements. We show that our method is able to retrieve a relatively complete ( $C \sim 85\%$ ) and pure ( $P \sim 90\%$ ) sample of Ly $\alpha$ -emitting quasars for  $L_{\text{Ly}\alpha} > 10^{44} \text{ erg s}^{-1}$ . In order to obtain corrections for the LF estimation, and assess the accuracy of our selection method, we produced mock catalogs of  $0 < z < 4.3$  quasars and galaxies that mimic our target population and their main contaminants. Our results show a clear evolution of the Ly $\alpha$  and UV LFs, with a declining tendency in the number density of quasars toward increasing redshifts. In addition, the faint-end power-law slope of the Ly $\alpha$  LF becomes steeper with redshift, suggesting that the number density of Ly $\alpha$ -bright quasars declines faster than that of fainter emitters. By integrating the Ly $\alpha$  LF, we find that the total Ly $\alpha$  emitted by bright quasars per unit volume rapidly declines with increasing redshift, being subdominant to that of star-forming galaxies by several orders of magnitude by  $z \sim 4$ . Finally, we stack the NB pseudo-spectra of a visually selected “golden sample” of 591 quasars to obtain photometric composite SEDs in bins of redshift, enabling us to measure the mean intergalactic medium absorption using the Lyman- $\alpha$  forest as a function of redshift, yielding results consistent with previous spectroscopic determinations.

**Key words.** line: identification – methods: observational – galaxies: high-redshift – galaxies: luminosity function, mass function – quasars: emission lines

## 1. Introduction

Active galactic nuclei (AGN) are extremely energetic events powered by efficient accretion onto super massive black holes (SMBH) in the cores of galaxies. These objects are among

the most luminous in the Universe across the full electromagnetic spectrum (e.g., Shen et al. 2020), and are easily detected at very high redshift by ground-based observatories ( $z > 6$ ; e.g., White et al. 2003; Jiang et al. 2016; Wang et al. 2021). Recent studies have confirmed that most massive galaxies host a SMBH in their center (e.g., Askar et al. 2022; Tremmel et al. 2024), and multiple lines of evidence have been found that these SMBH

\* Corresponding authors; alberto.torralba@uv.es,  
p.renard.guiral@gmail.com

coevolve with the bulge of the host galaxy (e.g., Yang et al. 2019; Mountrichas 2023). Additionally, most galaxy formation models require feedback from AGNs to prevent the interstellar gas from overcooling, quenching the star formation (e.g., Dubois et al. 2016; Davé et al. 2019; Ward et al. 2022) – otherwise, the models predict an excessively large number of luminous galaxies, in contradiction with observations (Fabian 2012). Hence, investigation of the properties of the AGN population and how it evolves with cosmic time is key in order to fully understand the bigger picture of galaxy evolution.

The luminosity function (LF) of AGNs, defined as the number density of these objects per comoving volume as a function of luminosity, evolves with cosmic time (e.g., Kulkarni et al. 2019; Shen et al. 2020), both in shape and normalization: the-faint end of the bolometric AGN LF peaks at  $z \approx 2-2.5$ , coinciding with the peak of star formation history (e.g., Madau & Dickinson 2014) whilst the bright-end peaks at a slightly earlier cosmic time of  $z \approx 3$ , and evolves slower toward the lowest redshifts, in what is often called cosmic downsizing (e.g., Hasinger et al. 2005; Hopkins et al. 2007; Croom et al. 2009; Fanidakis et al. 2012).

The Lyman- $\alpha$  ( $L_{\text{Ly}\alpha}$ ) emission-line is typically the most luminous in the rest-frame ultraviolet (UV) spectra of quasars (QSOs; e.g., Vanden Berk et al. 2001; Ivashchenko et al. 2014; Harris et al. 2016), and star forming galaxies (SFGs; e.g., Partridge & Peebles 1967; Pritchett 1994; Nakajima et al. 2018). The observed wavelength of  $L_{\text{Ly}\alpha}$  is redshifted into the optical range at  $z \sim 2-6$ , allowing us to detect this line in ground-based surveys. Sometimes, secure identification of  $L_{\text{Ly}\alpha}$  is possible even without explicit detection of the spectral continuum (e.g., Bacon et al. 2015). On the one hand, the faint end of the  $L_{\text{Ly}\alpha}$  LF is dominated by a population of typically small, young metal- and dust-poor SFGs ( $L_{\text{Ly}\alpha} \lesssim 10^{43.5} \text{ erg s}^{-1}$ ; e.g., Guaita et al. 2010; Drake et al. 2017; Arrabal Haro et al. 2020; Santos et al. 2020). The  $L_{\text{Ly}\alpha}$  LF deviates from a typical Schechter shape for  $L_{\text{Ly}\alpha} \gtrsim 10^{43.5} \text{ erg s}^{-1}$ , and the detection of X-ray counterparts of the sources at this range of  $L_{\text{Ly}\alpha}$  luminosities suggests that most of them are powered, at least partially, by AGNs (Sobral et al. 2018; Matthee et al. 2017). On the other hand, the brightest end of the  $L_{\text{Ly}\alpha}$  LF ( $L_{\text{Ly}\alpha} > 10^{43.5} \text{ erg s}^{-1}$ ) is dominated by QSOs (e.g., Calhau et al. 2020; Spinoso et al. 2020). The  $L_{\text{Ly}\alpha}$  LF in the whole range is well described by a double Schechter function (Schechter 1976). The intermediate  $L_{\text{Ly}\alpha}$  luminosity regime ( $L_{\text{Ly}\alpha} \sim 10^{43.5} \text{ erg s}^{-1}$ ) is particularly interesting, because it might contain objects where the  $L_{\text{Ly}\alpha}$  emission from the AGN and star-formation activity is similarly relevant (e.g., Sobral et al. 2018). While the  $L_{\text{Ly}\alpha}$  LF of SFGs has been widely explored in the literature (e.g., Cowie & Hu 1998; Hu et al. 1998; Martin & Sawicki 2004; Malhotra & Rhoads 2004; Dawson et al. 2007; Ouchi et al. 2008; Blanc et al. 2011; Adams et al. 2011; Bacon et al. 2015; Karman et al. 2015; Drake et al. 2017; Santos et al. 2016; Konno et al. 2018; de La Vieuville et al. 2019; Hu et al. 2019; Khusanova et al. 2020; Ono et al. 2021; Santos et al. 2021; Morales et al. 2021; Wold et al. 2022; Xu et al. 2023; Thai et al. 2023), the AGN-dominated bright-end of the  $L_{\text{Ly}\alpha}$  LF has only been constrained in recent years (Spinoso et al. 2020; Zhang et al. 2021; Liu et al. 2022; Torralba-Torregrosa et al. 2023).

In the same manner as the  $L_{\text{Ly}\alpha}$  LF, the UV LF is dominated by SFGs at the faint end ( $M_{\text{UV}} \gtrsim -23$ ), and AGN at the bright end ( $M_{\text{UV}} \lesssim -23$ ; e.g., Adams et al. 2023). Contrary to the  $L_{\text{Ly}\alpha}$  LF, the bright end of the UV AGN LF has been broadly explored (e.g., Glikman et al. 2011; Ross et al. 2013; McGreer et al.

2013; Palanque-Delabrouille et al. 2016; Jiang et al. 2016; Yang et al. 2016; Kulkarni et al. 2019; Akiyama et al. 2018; Matsuoka et al. 2018; Schindler et al. 2019; Niida et al. 2020; Pan et al. 2022). However, the faintest end of the AGN UV LF is still uncertain. In particular, a population of faint dust-obscured AGNs has been unveiled by the most recent JWST observations (e.g., Labbé et al. 2023; Matthee et al. 2024; Kokorev et al. 2024; Kocevski et al. 2024; Greene et al. 2024; Akins et al. 2024), whose number density appears to be larger by more than one order of magnitude than what is expected when extrapolating the bright end of the AGN UV LF. The connection between this population of faint dusty AGNs and bright QSOs is still unclear, and the determination of the AGN LF in the intermediate-luminosity regime ( $M_{\text{UV}}$  between  $-24$  and  $-20$ ) is crucial in order to solve this problem. A global determination of the AGN LF is key to determining the role of this population in cosmic reionization. Hydrodynamical and radiative transfer simulations suggest that the contribution of AGNs to hydrogen reionization is subdominant (e.g., Dayal et al. 2020; Trebitsch et al. 2021). This is supported by the rapid decline in the number of bright quasars at high redshifts (e.g., Kulkarni et al. 2019). However, the role of AGNs can be determinant in helium reionization, which happened at later cosmic time (e.g., Compostella et al. 2014; Worseck et al. 2016).

In this context, in the present work we seek to identify and characterize the population of quasars with bright ( $L_{\text{Ly}\alpha} > 10^{43.5} \text{ erg s}^{-1}$ )  $L_{\text{Ly}\alpha}$  emission using the NB data from the Physics of the Accelerating Universe Survey (PAUS). Multi-NB surveys such as this latter are extremely suitable for this task. On the one hand, the continuous coverage of a great part of the optical range with NB photometry yields a low-resolution spectrum for every source in the field of view (FoV) without any target preselection, allowing very complete samples if carefully selected. On the other hand, it becomes possible to probe a wide and continuous range of redshift, which is not the case for surveys that only use a few NBs placed in strategic wavelengths. In addition, the availability of pseudo-spectra can be used to obtain precise photometric redshifts (Martí et al. 2014; Eriksen et al. 2020; Alarcon et al. 2021; Hernán-Caballero et al. 2021; Laur et al. 2022; Navarro-Gironés et al. 2023; Hernán-Caballero et al. 2024).

In Torralba-Torregrosa et al. (2023, hereafter TT23), we presented a method to select  $L_{\text{Ly}\alpha}$  emitters in the miniJPAS (Bonoli et al. 2021) and J-NEP (Hernán-Caballero et al. 2023) multi-NB data. In the present work, we adapt and improve the selection method to the PAUS data, and use it to estimate the UV and  $L_{\text{Ly}\alpha}$  LFs of QSOs selected through  $L_{\text{Ly}\alpha}$  emission. We also use the obtained sample of  $L_{\text{Ly}\alpha}$ -selected QSOs to estimate the mean absorption by the intergalactic medium (IGM) due to the  $L_{\text{Ly}\alpha}$  forest as a function of redshift (e.g., Faucher-Giguère et al. 2008; Becker et al. 2013; Inoue et al. 2014).

This paper is structured as follows. In Sect. 2 we describe the observational data obtained during the PAU Survey and employed throughout this work. In Sect. 3 we first describe the mock catalogs, which are used to test and validate our selection method; we define our  $L_{\text{Ly}\alpha}$  selection method, and the procedures we use to estimate the photometric redshifts of the sources and the  $L_{\text{Ly}\alpha}$  luminosity, and to obtain the  $L_{\text{Ly}\alpha}$  and UV LFs. In Sect. 4 we present the  $L_{\text{Ly}\alpha}$ -selected sample, and produce photometric composite QSO spectra in bins of redshift. We present the  $L_{\text{Ly}\alpha}$  and UV LFs, and their integrated values as a function of redshift. In Sect. 5 we compare our measured  $L_{\text{Ly}\alpha}$  and UV LFs with previous determinations in the literature, and discuss the evolution

of these with redshift. We also comment on the efficiency of our selection method, comparing with other AGN samples in the literature, and cross-matching our catalog with spectroscopic surveys. Finally, Sect. 6 summarizes the contents of this work.

Throughout this work, we use a  $\Lambda$ CDM cosmology as described by [Planck Collaboration VI \(2020\)](#), with  $\Omega_\Lambda = 0.69$ ,  $\Omega_M = 0.31$ , and  $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , unless specified otherwise. All the magnitudes are given in the AB system ([Oke & Gunn 1983](#)).

## 2. Observations: The PAU Survey

The Physics of the Accelerating Universe Survey (PAUS) is a wide-field multi-NB survey carried out by the PAUCam instrument ([Padilla et al. 2019](#)), mounted at the 4.2 m *William Herschel* Telescope (WHT) at the Roque de los Muchachos Observatory in the Canary Islands. PAUCam is an 18-detector camera, with a FoV of  $\sim 1^\circ$  in diameter, designed to cover a wide area of the sky making use of a filter set composed of 40 NBs (FWHM  $\sim 115 \text{ \AA}$ ). The unique set of NB filters provides continuous wavelength coverage over observed wavelengths of 4500–8500  $\text{\AA}$ , allowing a low-resolution pseudo-spectrum to be obtained for every source in the FoV. The detection of diffuse Ly $\alpha$  emission in PAUS has already been evaluated in [Renard et al. \(2021\)](#), and other prominent spectral features have already been studied with PAUS data, such as emission line ratios ([Alarcon et al. 2021](#)), or the D4000 spectral break ([Renard et al. 2022](#)).

PAUS targeted the CFHTLS ([Cuillandre et al. 2012](#)) W1, W3, and W4 fields; the GAMA G09 field ([Driver et al. 2022](#)); and the COSMOS field ([Scoville et al. 2007](#)). In this work, we use the data of W1, W3, and G09 (hereafter dubbed W2). In total, these three fields are covered with the full set of 40 filters in an effective area of 35.68  $\text{deg}^2$  (after masking). For PAUS catalogs, all objects with magnitudes brighter than  $i = 23$  are selected; at fainter magnitudes, the photometry S/N is too low to reliably measure galaxy redshifts ([Alarcon et al. 2021](#); [Navarro-Gironés et al. 2023](#)).

The standard PAUS photometry is computed using forced photometry ([Serrano et al. 2023](#)). For each object and narrow-band, its fluxes are separately measured on each single-exposure image. Three exposures are taken per pointing, though the number of exposures on individual objects may vary if they are positioned near the edges of the tiling, due to overlaps and dithering. The flux apertures are ellipses that contain 62.5% of the flux, according to their Sérsic profiles convolved with the seeing of each exposure; the total flux for a given object and narrow band is the inverse-variance weighted average of all exposures. This method results in a custom aperture for each object, fitted to their size and morphology, which in turn yields a higher S/N than standard fixed apertures ([Serrano et al. 2023](#)). For this work, we computed custom photometric runs of W1, W2 and W3 utilizing the same data reduction outlined in [Serrano et al. \(2023\)](#), but using fixed circular apertures of diameter  $3''$  instead. This fixed aperture was chosen in order to be consistent with the photometry used in [TT23](#).

The positions of all objects, their  $i$  magnitudes and morphological parameters are extracted from broad-band reference catalogs; no detections are performed on PAUS narrow-band images. The reference catalog used for the W1 and W3 fields is the CFHTLS data release from [Heymans et al. \(2012\)](#), and for the W2 (GAMA G09) field it is the fourth KiDS data release ([Kuijken et al. 2019](#)).

## 3. Methods

In this section we describe the methodology employed to select high-redshift QSOs in the PAUS wide fields, and the estimation of the Ly $\alpha$  and UV luminosity functions. We first describe the mock catalogs developed to mimic our observations, which we use to test and assess the effectiveness of our methods (Sect. 3.1). Next, we describe our target selection method, based on Ly $\alpha$  emission detection (Sect. 3.2), we characterize the purity and completeness of our method (Sect. 3.3) and describe the procedure to estimate the Ly $\alpha$  and UV luminosity functions (Sect. 3.4).

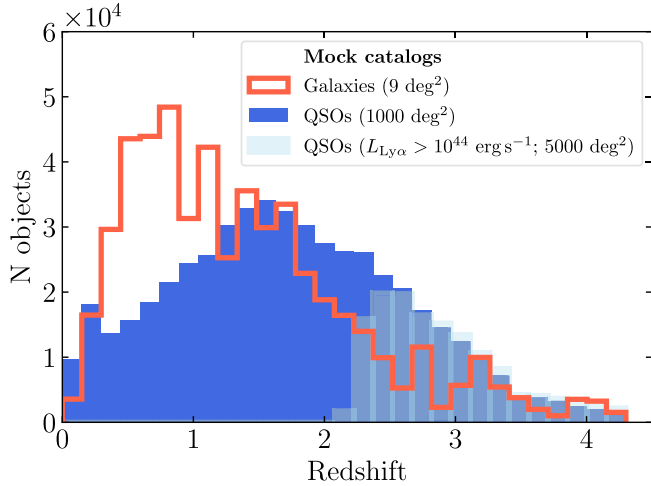
### 3.1. Mock catalogs

We generate mock catalogs and photometric NB pseudo-spectra of Ly $\alpha$  emitting QSOs (our target sample) and their common contaminants (low- $z$  QSOs and galaxies), with the aim of characterizing the effectiveness of our selection method. The mocks are employed to train machine learning algorithms to improve the selection function (Sect. 3.2) and to measure accurate photometric redshifts (Sect. 3.2.6); to estimate the purity and completeness of the selected sample (Sect. 3.3) and to obtain corrections for the estimation of the Ly $\alpha$  and UV luminosity functions (Sect. 3.4).

#### 3.1.1. QSO mock catalog

The mock catalog of the QSO population at  $0 < z < 4.5$  is generated following the same procedure as in [TT23](#), which at the same time is based on the method presented in [Queiroz et al. \(2023\)](#) for a QSO mock population in the J-PAS survey. We select spectra in this redshift range from the SDSS DR16 SuperSetQ ([Lyke et al. 2020](#)) with good median signal-to-noise ratio over all pixels ( $\text{SN\_MEDIAN\_ALL} > 0$ ), a measured redshift with a high level of confidence ( $\text{ZWARNING} = 0$ ), and classified as QSO ( $\text{IS\_QSO\_FINAL} = 1$ ). Next, we sample values over a grid of redshift and  $r$ -band magnitude from the luminosity functions of [Palanque-Delabrouille et al. \(2016\)](#) (PLE+LEDE model). We look for the closest objects in redshift within the SDSS DR16 SuperSetQ subsample, and apply the corresponding multiplicative factors to the observed wavelength array and fluxes to match the sampled values of  $(z, r)$ . In [TT23](#), we show that this procedure yields a population of QSOs that, for  $z > 2$ , is in good agreement with the Ly $\alpha$  LF measured in the literature (see Fig. 2 in [TT23](#)). We produce a QSO mock with a size equivalent to 1000  $\text{deg}^2$ , and a five times larger area for objects with  $L_{\text{Ly}\alpha} > 10^{44} \text{ erg s}^{-1}$  to overcome the limitations due to the low number of objects at such high luminosities. The sizes of the simulated areas are chosen as arbitrarily large values sufficient to obtain smooth distributions in the mock properties (for more details about the QSO mock, see [TT23](#)). The number of objects in these mocks as a function of redshift is shown in Fig. 1.

As a final step, we compute synthetic photometry of all the selected objects using the transmission curves of all the PAUS filters to obtain pseudo-spectra. Since all the SDSS spectra used to generate the mocks are selected to have good S/N, the error of our synthetic photometry is negligible compared to the error of real PAUS observations. We model the PAUS photometric errors separately. For this, we model the uncertainties as a function of photometric magnitude with an exponential (see [TT23](#)), fitting the exponential parameters for every PAUS filter in each one of the W1, W2 and W3 fields. We add random Gaussian perturbations to the synthetic fluxes of the mock according to this model to obtain the final mock QSO catalog.



**Fig. 1.** Redshift distribution of galaxies within the QSO and low- $z$  galaxy mocks. The different mock populations were simulated in different areas: 9, 1000, and 5000  $\text{deg}^2$  for the low- $z$  galaxy, QSOs, and high  $L_{\text{Ly}\alpha}$  QSO mocks, respectively. In all mocks, we applied conservative magnitude cuts of  $r < 25$ . Despite the much larger area used in the QSO mocks, the low- $z$  galaxies are significantly more numerous.

### 3.1.2. Low-redshift galaxy mock catalog

Among the main contaminants in the search for Ly $\alpha$  emission there are nebular emission-lines associated with star-formation such as H $\alpha$  ( $\lambda$  6565 Å), H $\beta$  ( $\lambda$  4861 Å), [O III] ( $\lambda$  4959, 5007 Å) or [O II] ( $\lambda$  3727, 3729 Å) emitted by low- $z$  galaxies (i.e.,  $z < 2$ , in this context). On top of their eventual bright line-emission, low- $z$  galaxies may contaminate our LAE sample due to random fluctuations in their NB fluxes. This source of contamination may be relevant because low- $z$  galaxies account for the bulk of the objects in the parent photometric catalog. Despite this, TT23 found little contamination of nebular emission lines from  $z < 2$  galaxies, with the minor exception of a few extreme [O II] emitters.

In order to assess the impact of low- $z$  galaxies as contaminants in our samples, we make use of a galaxy mock, built by using the semi-analytic model (SAM) L-Galaxies on top of the dark-matter merger trees of the Millennium N-body dark matter (DM)-only simulation (Springel et al. 2005). In particular, the version of L-Galaxies we use is derived from the public release of Henriques et al. (2015), with the inclusion of detailed models for the formation, mass-growth and dynamical evolution of SMBHs recently presented in Izquierdo-Villalba et al. (2020, 2022) and Spinoso et al. (2023).

On top of being one of the most detailed SAMs in the literature to describe the evolution of galaxies and SMBHs, L-Galaxies includes the possibility to produce photometric mock-data on-the-fly. Indeed, Izquierdo-Villalba et al. (2019) presented a methodology to simulate an observed light-cone for a generic multiband photometric survey, showing the example of the J-PLUS survey (Cenarro et al. 2019). In particular, their method relies on simulating the continuum emission of galactic stellar populations by means of a galaxy-templates grid which depends on metallicity, stellar age and a given choice for the stellar initial mass function (Izquierdo-Villalba et al. 2019 opted for a standard Chabrier IMF; Chabrier 2003). On top of the stellar continua, nine different nebular emission lines<sup>1</sup> from star-

<sup>1</sup> Namely: Ly $\alpha$  (1215.67 Å), H $\beta$  (4861 Å), H $\alpha$  (6563 Å), [O II] (3727 Å, 3729 Å), [O III] (4959 Å, 5007 Å), [Ne III] (3870 Å), [O I] (6300 Å), [N II] (6548 Å, 6583 Å), and [S II] (6717 Å, 6731 Å).

forming regions are simulated by using the method presented in Orsi et al. (2014), where the emission line flux is obtained as a function of stellar age, gas density, metallicity and ionization parameter (see Izquierdo-Villalba et al. 2019, for further details on the implementation of the emission-lines model in L-Galaxies).

In order to build our low- $z$  galaxy mock we follow the methodology outlined above, adapting it to the PAUS photometric filter-set and survey volume. In particular, we simulate a  $3 \times 3 \text{ deg} = 9 \text{ deg}^2$  sky patch; that is, roughly one-fourth of the effective PAUS area we use (see Sect. 3.2.1). The choice of this area is mainly limited by computation time for the lightcone. Despite the much smaller area as compared to the QSO mock, the number of galaxies in the lightcone is similar to the number of objects in the QSO mock (see Fig. 1). We apply a conservative magnitude cut (i.e.,  $i < 25$ ) to the photometry of galaxies within the light-cone. Given the effective depth of PAUS data, this magnitude cut allows us to significantly reduce the size of our low- $z$  galaxy mock data without compromising their completeness. The PAUS photometric errors are simulated as described in Sect. 3.1.1 for the QSO mock. Our final mock catalog contains  $\sim 5 \times 10^5$  galaxies with  $i < 25$  at redshifts  $0 < z \lesssim 6$  (median  $z = 1.2$ ) (see Fig. 1).

## 3.2. Lyman- $\alpha$ selection method

In this section we describe the selection method used to obtain a Ly $\alpha$ -selected sample of QSOs in the footprint of PAUS.

### 3.2.1. Parent sample

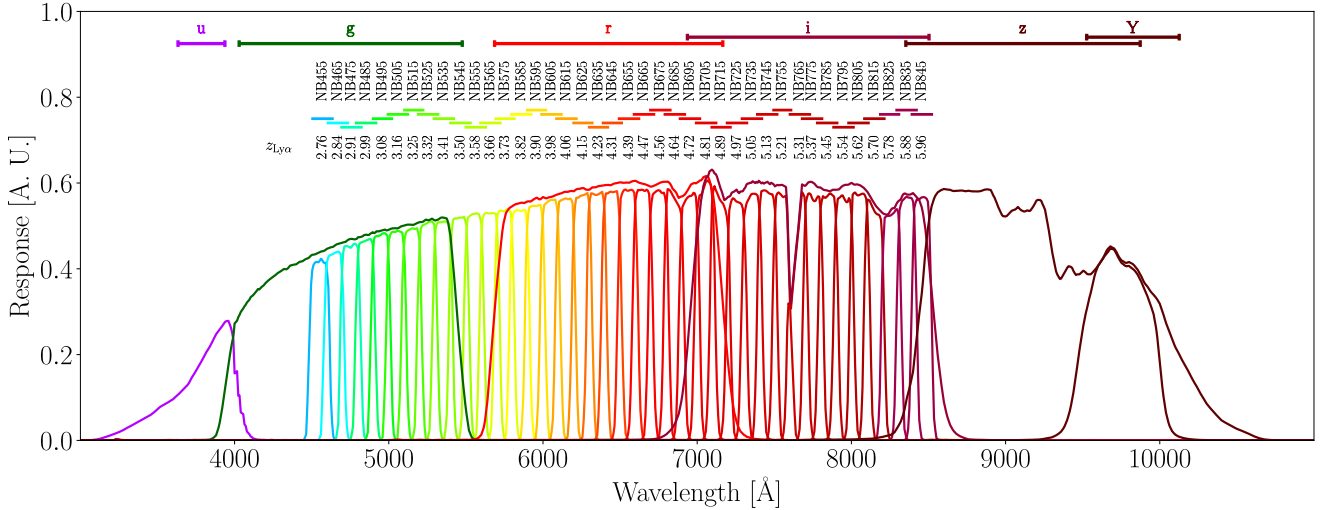
We use the diameter 3'' aperture photometry catalogs of the PAUS wide fields W1, W2 and W3 as discussed in Sect. 2. For our parent sample, we use only the regions of these fields where there is coverage for all 40 PAUS NBs. The reference catalogs used by these fields are taken from surveys with deeper broad-band imaging and better PSF than the average of PAUS NBs. Due to this, some sources from the reference catalogs are blended in the PAUS NB images, leading to contamination in the fixed aperture photometry. Hence, we remove objects with neighbours with a separation closer than 3''. This cut removes 8% of the objects across the W1, W2, and W3 catalogs, and is therefore taken into consideration in the computation of the LF.

After removing potentially blended sources, and applying the PAUS field masks for the area covered by all 40 NB filters, we are left with a parent sample containing 1 022 196 objects (239 312, 267 393 and 515 491 in W1, W2 and W3, respectively) in a total area of 35.7  $\text{deg}^2$  (effective area after masking).

### 3.2.2. NB flux-excess selection

We search for Ly $\alpha$  emission in each of the PAUS narrow bands through a NB flux-excess selection. Each narrow band probes Ly $\alpha$  lines at redshifts within  $\delta z = \pm 0.06$  around the central redshifts listed in Table A.1, and displayed in Fig. 2. This redshift interval is defined by the FWHM of the NB filters. A detailed description of the LAE candidate selection method can be found in TT23. In the following sections we summarize the method, which was adapted for the particularities of PAUS data.

In the first place, we look for tentative Ly $\alpha$  detections in all the NBs. For each NB, an estimation of the continuum flux is obtained by averaging the photometric fluxes of the 5 closest NBs to the one of interest, at bluer and redder wavelengths,



**Fig. 2.** Transmission curves of PAUS NBs and BBs of the reference catalogs. On top of the transmission curves, horizontal lines are displayed showing the FWHM of each filter. We show the Ly $\alpha$  redshift corresponding to the pivot wavelength of each NB. As shown in Table A.1, only NB455–NB755 are used in this work, as we obtained no reliable detections in NBs of higher wavelengths.

ignoring both directly adjacent NBs to avoid contamination from the tail of broad Ly $\alpha$  lines (see e.g., Greig et al. 2017). That makes a total of 10 NB filters to compute the continuum level. In addition, the flux of the NBs blueward from Ly $\alpha$  is affected by the Ly $\alpha$  forest. The mean absorption of the Ly $\alpha$  forest is corrected in the continuum estimation using the curve for the mean IGM transmission in Faucher-Giguère et al. (2008). After the estimation of the continuum flux level of every source in our parent sample, we look for NBs with S/N > 6 and a 3 $\sigma$  flux excess over the continuum:

$$f_{\text{NB}}^{\lambda} - f_{\text{cont}}^{\lambda} > 3 \sqrt{\sigma_{\text{NB}}^2 + \sigma_{\text{cont}}^2}. \quad (1)$$

We also impose that this excess is compatible with a Ly $\alpha$  line with  $\text{EW}_0 \geq \text{EW}_0^{\text{min}} = 20 \text{ \AA}$  by applying

$$\frac{f_{\text{NB}}^{\lambda}}{f_{\text{cont}}^{\lambda}} > 1 + \frac{(1 + z_{\text{NB}}) \cdot \text{EW}_0^{\text{min}}}{\text{FWHM}_{\text{NB}}}, \quad (2)$$

where  $z_{\text{NB}}$  is the Ly $\alpha$  redshift corresponding to the target NB. (see TT23, also e.g., Santos et al. 2016; Sobral et al. 2018; Spinoso et al. 2020). If two or more contiguous NBs meet these criteria, the Ly $\alpha$  detection is assigned to that of higher flux.

After the preliminary Ly $\alpha$  detection, we similarly search for other lines in the selected sources. For this secondary line search, following the optimal search criteria in TT23, we look for NBs with >3 $\sigma$  flux over continuum excess and  $\text{EW}_{\text{obs}} > 100 \text{ \AA}$ . These emission lines are searched using a continuum estimate without the Ly $\alpha$  forest transmission correction. The algorithm checks the secondary line detections and discards candidates with lines not compatible with the expected position of the most prominent QSO lines other than Ly $\alpha$ ; in particular Ly $\beta$ +O VI, C IV, C III] and Mg II. Figure 3 shows four hand-picked examples of selected candidates using the procedure described above. The first three panels show correctly selected QSOs with coinciding spectroscopic Ly $\alpha$  redshifts of their SDSS counterparts. In these three examples, we also detect the secondary C IV line, which ensures a correct determination of the redshift. The bottom panel shows an example of the most common contaminant: a confusion between Ly $\alpha$  and a lower redshift ( $z = 2.3$ ) C IV emission line.

### 3.2.3. Artificial neural network classifier

In TT23 a visual inspection was performed through all the candidates in order to remove obvious contaminants such as low- $z$  galaxies with extended morphology, or NBs affected by cosmic rays and other artifacts. In this work, we trained a multilayer perceptron artificial neural network (ANN) model in order to identify secure candidates and discard contaminants. This is motivated by the notable increase in the size of the candidate sample with respect to that in TT23, and the potential application of the method to even larger datasets in the future.

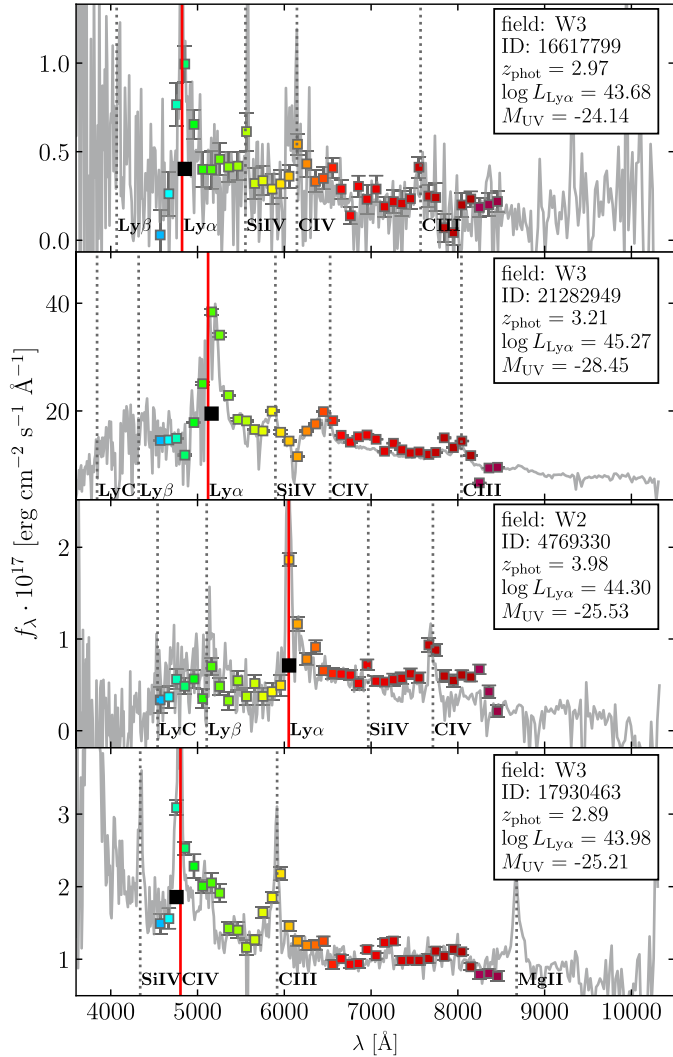
In order to train the ANN, we generate training and test sets from the subsets of objects in our mocks selected by NB flux excess (see Sect. 3.2.2). Each set contains equal number of the three classes to consider: (1) QSO with the correct narrow-band redshift, (2) QSO with wrong narrow-band redshift and (3) low- $z$  galaxy. The first of these three classes corresponds to the true positive LAE detections, while the other two represent the main contaminants.

The input for the ANN classifier is composed of 46 features:

- All 40 NB fluxes. Normalized to the maximum NB flux of each object.
- $g$ ,  $r$ ,  $i$ , and  $z$  BB fluxes. Normalized to the maximum BB flux. The band  $u$  is omitted for being partially out of the SDSS range used to create the mocks.
- $r_{\text{synth}}$ . Synthetic broad-band magnitude computed from the NB whose wavelength coverage reproduces the classical  $r$  band ( $\lambda \approx 5750\text{--}7000 \text{ \AA}$ ). Normalized with a factor 1/30 (so every value is <1).
- $\text{NB}_{\text{Ly}\alpha}$ : the NB of the Ly $\alpha$  detection, numbered from 0 to 39. Normalized with a factor 1/39.

The NB and BB arrays of fluxes are normalized separately since the BB data is obtained from reference catalogs external to PAUS (see Sect. 2), and their fluxes might not be directly comparable to PAUS 3'' aperture NB fluxes. The  $r_{\text{synth}}$  feature gives information about the apparent magnitude of the object, and  $\text{NB}_{\text{Ly}\alpha}$  establishes an initial guess of the redshift of the source. A synthetic composite  $r_{\text{synth}}$  broad-band is used instead of  $r$  from the reference catalogs for consistency.

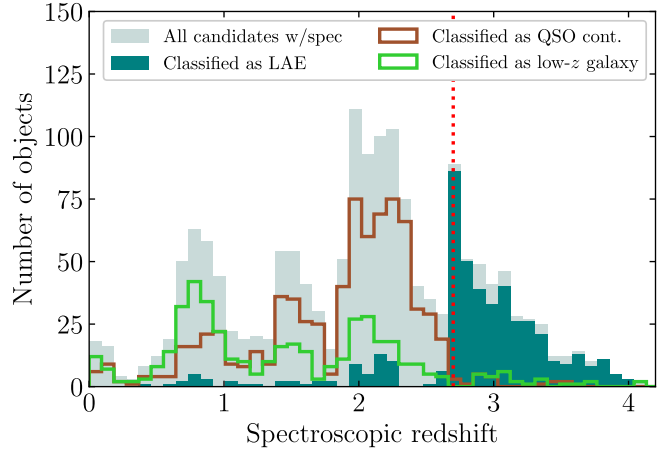
We perform a search for optimal parameters of the ANN over a wide parameter space. The optimal parameters obtained in the



**Fig. 3.** Four examples of candidates selected by our pipeline. The colored squares represent the NB fluxes of PAUS. We compare with the SDSS spectrum in gray (resampled by a factor 10 to reduce visual noise). We mark the most important QSO emission lines (dotted gray vertical lines) and the predicted observed wavelength of Ly $\alpha$  by our selection pipeline (vertical red line). The estimation of the continuum level is marked with a black square. The first three panels are examples of correctly selected Ly $\alpha$  emitters. The last panel is an example of a confusion between Ly $\alpha$  and C IV ( $z_{\text{CIV}} = 2.1$ ), the most common source of contamination.

grid-search, and more details about the ANN architecture can be found in Appendix B. As shown in Fig. B.1, the classifier yields a  $\sim 90\%$  pure sample of QSO LAEs over the test set ( $\sim 95\%$  for  $L_{\text{Ly}\alpha} > 10^{44} \text{ erg s}^{-1}$ ), with 7% (2%) contamination from QSOs (galaxies) at lower redshifts. The completeness of the sample classified as QSO LAE is 96% and 98% in the test set, for the general sample and for  $L_{\text{Ly}\alpha} > 10^{44} \text{ erg s}^{-1}$ , respectively.

We also determine the accuracy of the ANN classifier with our spectroscopic subsample (presented in Sect. 5.4 below). In Fig. 4 we show the classification over this subsample. For  $z \geq 2.75$ , our target redshift range (see Table A.1), the classifier correctly identifies as LAEs  $\sim 90\%$  of the objects, in line with the completeness estimated in the test set. For lower redshift,  $z \leq 2.75$  (i.e., the range of the contaminants), the algorithm successfully identifies most of the objects as contaminants. The confusion between contaminant QSO and galaxies is irrelevant in this work, since we are only interested in the QSO LAE sub-



**Fig. 4.** Classification of the candidates with a spectroscopic counterpart by our ANN algorithm. The solid light-teal histogram represents the full spectroscopic sample, and the solid dark blue the subsample classified as LAE by the NN. The empty brown and green histograms represent the candidates classified as contaminant QSOs and low- $z$  galaxies, respectively. For our target redshift range ( $z \geq 2.75$ ), the algorithm correctly classifies most objects ( $\sim 90\%$  completeness). The red dotted vertical line marks the minimum redshift targeted by the NB with the lowest wavelength ( $z = 2.7$ , NB455).

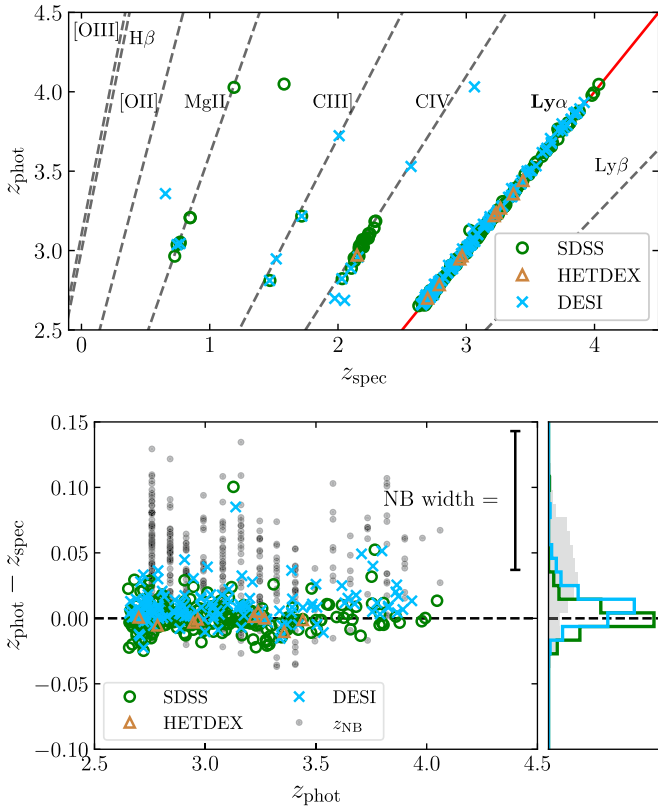
sample. The purity derived from this result is  $\sim 86\%$  ( $>90\%$  for  $L_{\text{Ly}\alpha} > 10^{44} \text{ erg s}^{-1}$ ), again in line with that of the test set.

### 3.2.4. Visual inspection

The mock data used to train the ML model described in Sect. 3.2.3 only contains objects with clean synthetic photometry, performed over either simulated SEDs (in the case of the low- $z$  galaxy mock) or spectra with no warnings raised by the SDSS pipelines (in the case of the QSO mocks). Hence, the training set does not account for objects affected by instrumental artifacts, cosmic rays or photometry calibration problems, among other issues that can affect the NB fluxes leading to a false detection. For this reason, despite the good performance shown by the ML model, a visual inspection is still necessary in order to catch objects misclassified by the algorithm due to unexpected features. Nonetheless, the ML classification significantly reduced the candidate sample for the visual analysis, from an initial sample of 4356 objects to 999 candidates.

The NB photometry of some objects is clearly affected by the extended emission of a nearby object, leading to unusually high NB fluxes that are mistaken by Ly $\alpha$  emission lines. In some other objects, the images were affected by cosmic rays, instrumental artifacts or similar effects that alter the NB fluxes. Finally, some images of the PAUS fields are affected by calibration issues, resulting in unusually large zero-points. The zero-points are computed with a cross-match of bright SDSS stars on a per-CCD basis (Serrano et al. 2023, each CCD zero-point being the median of the reference stars). Hence, flux underestimation of reference stars (e.g., due to scatter-light or cross-talk between CCDs) could result on abnormally high zero-points for a few exposures. This issue artificially enhances the measured flux of some NBs in some regions, leading to wrong Ly $\alpha$  line detections. All these problems cannot be trivially reproduced in our mocks, and thus the affected objects must be removed from the selected sample via visual classification.

We visually inspected the initial sample of 999 QSO LAE candidates selected by our method, and agreed on the final



**Fig. 5.** Comparison between photometric and spectroscopic redshifts. Top: Comparison between photometric and spectroscopic redshifts of the candidates in our sample with counterparts in SDSS DR18 (green circles) HETDEX (orange triangles) and DESI DR1 (blue crosses). The straight lines mark the confusion between  $\text{Ly}\alpha$  and other bright QSO lines. Thus, objects lying on the red line are true positive detections of LAEs. Bottom: Difference between the photometric and spectroscopic redshifts of the objects shown in the top panel. The black dots represent the same objects using the photometric redshift of the selection NB, before applying the machine-learning algorithm presented in Sect. 3.2.6. The colored dots represent the same objects after the random forest redshift correction, for each spectroscopic subsample. The bottom right panel shows normalized histograms, the result of collapsing the left panel in the horizontal axis. The HETDEX histogram is not shown because of poor statistics.

classification. The inspection yielded a “golden sample” of 591 (59%) objects selected as true positive QSO LAEs at the correct redshift. On the other hand, a non-negligible amount of 102 (10%) objects were removed from the final sample due to issues with their pseudo-spectra or images. Some of these removed objects showed contamination by cosmic rays in their selected NB, evident instrumental artifacts, extended emission from nearby extended objects or clear issues with the photometric calibration that led to unusual pseudo-spectra. The remaining 316 (31%) objects are left as “uncertain”, not being classified as true positives nor as contaminants.

### 3.2.5. High- $z$ sample

In addition to the search of candidates in the interval  $2.7 < z < 4.2$ , we look for objects at higher redshifts. However, since our machine learning algorithm was trained using SDSS data, it is limited to the above mentioned redshift range, given that the SDSS sample becomes too scarce for higher redshifts. We perform a candidate search for  $z > 4.2$ , without applying the ANN

classification, and inspect all selected objects to identify secure  $\text{Ly}\alpha$  emitting QSOs. This search produced a sample of 22 visually selected QSOs ( $\sim 0.6 \text{ deg}^{-1}$ ) at  $4.2 \lesssim z \lesssim 5.3$  (see Table A.1), from an initial sample of 549 high- $z$  candidates.

### 3.2.6. NB redshift correction

Each NB can probe an observed  $\text{Ly}\alpha$  wavelength in an interval defined by the width of the transmission curve. The mean redshift probed by the PAUS NBs used in this work are summarized in Table A.1. In addition, the full 40-NBs pseudo-spectra provide information that can be used to better constrain the redshifts of the detected sources. This is especially relevant for the correct assessment of the redshift of sources selected via  $\text{Ly}\alpha$  detection. The  $\text{Ly}\alpha$  emission-lines of these sources typically present broad profiles (width  $\sim 4500 \text{ km/s}$ ; Greig et al. 2016), and the absorption due to the  $\text{Ly}\alpha$  forest (for  $\lambda_{\text{obs}} < \lambda_{\text{obs}}^{\text{Ly}\alpha}$ ) typically induces a bias in the  $\text{Ly}\alpha$  redshift towards slightly higher values.

We improve the redshift accuracy of our selected objects using machine learning algorithms. In order to obtain train/test sets, we run our selection method on our mocks. Then we keep only those candidates whose difference between the NB redshift ( $z_{\text{NB}}$ ; i.e., the mean redshift of the selected NB) and the true redshift of the mock object ( $z_{\text{spec}}$ ) is smaller than the width of a NB in redshift space ( $\Delta z \lesssim 0.12$ ; see Table A.1). We train ANN and random forest (RF) regressors with the aim of retrieving the spectroscopic redshift for the objects in our test set. The input for the ML redshift regressors is the same as for the ANN classifier described in Sect. 3.2.3. We perform a search of optimal parameters for the ANN and RF algorithms over an wide grid. We find that the RF regressor greatly outperforms the ANN and is less affected by over-fitting for the optimal parameters of both.

Bottom panel of Figure 5 shows the difference between the redshift obtained by the RF and the spectroscopic redshift of all the selected objects in the PAUS catalogs with spectroscopic counterparts in SDSS, DESI and HETDEX. The initial NB redshifts are biased with respect to those of the spectroscopic counterparts, and present larger variance than the redshifts obtained by the RF: the bias and standard deviation are  $(\mu, \sigma)(\Delta z) = (0.04, 0.03)$  for the initial NB redshifts, and  $(\mu, \sigma)(\Delta z) = (0.003, 0.013)$  and  $(0.010, 0.012)$  for the redshifts produced by the RF, for the SDSS and DESI spectroscopic subsamples, respectively. The redshift bias of the DESI spectroscopic subsample is slightly larger than that of the SDSS sample. This can be explained by the fact that the mocks employed to train the RF algorithm are based on SDSS spectra. However, the RF estimation significantly improves the precision of the redshift measurements within the global spectroscopic sample.

### 3.2.7. $\text{Ly}\alpha$ luminosity correction

The luminosity of the selected  $\text{Ly}\alpha$  lines is estimated as

$$L_{\text{Ly}\alpha} \approx (f_{\text{NB}}^{\lambda} - f_{\text{cont}}^{\lambda}) \cdot \text{FWHM}_{\text{NB}} \cdot 4\pi d_L^2(z), \quad (3)$$

where  $f_{\text{NB}}^{\lambda}$  and  $f_{\text{cont}}^{\lambda}$  are the NB and continuum fluxes;  $\text{FWHM}_{\text{NB}}$  the full width at half maximum of the NB filter, and  $d_L$  the luminosity distance according to our cosmology. Several effects can alter the measurement of the  $\text{Ly}\alpha$  flux. In the first place, the estimation of  $f_{\text{cont}}$  is performed assuming a mean IGM transmission for the filters at shorter wavelengths than  $\text{Ly}\alpha$  (see Sect. 3.2), yet there can be significant variability in the  $\text{Ly}\alpha$  forest for individual sources, especially at the lowest redshift bins (e.g., Rollinde et al. 2013). Emission lines from QSOs typically show broad profiles, so the tails of the  $\text{Ly}\alpha$  profile may be partially

out of the coverage of a given NB. This effect is also enhanced if the intrinsic Ly $\alpha$  peak is not well aligned with the NB effective wavelength. In addition, in Eq. (3) a top-hat filter profile is assumed<sup>2</sup>. Finally, while most spectroscopic surveys are capable of disentangling the N V QSO line ( $\lambda 1240.81$  Å) from Ly $\alpha$ , this task is not possible with NB photometry, since the Ly $\alpha$  and N V are cannot be separated within the single flux measurements obtained with PAUS NBs. The flux of N V is typically fainter than Ly $\alpha$ , but it can significantly affect the Ly $\alpha$  measurement (see e.g., Vanden Berk et al. 2001; Selsing et al. 2016).

We empirically correct for these biases based on the comparison of our measured Ly $\alpha$  fluxes with the spectroscopic determinations of the SDSS DR16 counterpart subsample. For this, we determine the systematic offset between our flux measurements and those of SDSS as a function of  $L_{\text{Ly}\alpha}$ , and subtract it from the flux measurements of our general sample (up to  $\Delta \log_{10} L_{\text{Ly}\alpha} \sim 0.2$ ; see Fig. 4 in TT23).

### 3.3. Purity and completeness

Following TT23, we compute 2D maps of purity and number count corrections in order to estimate the Ly $\alpha$  LF. We apply our selection method (Sect. 3.2) to the mocks, and distribute the selected mock sample over a 2D grid of measured Ly $\alpha$  luminosity and synthetic  $r$  band magnitude (i.e.,  $L_{\text{Ly}\alpha}^{\text{obs}}$ ,  $r_{\text{synth}}$ ). We then compute the 2D purity as

$$P^{2\text{D}} = \frac{N_{\text{sel., true}}}{N_{\text{sel.}}}, \quad (4)$$

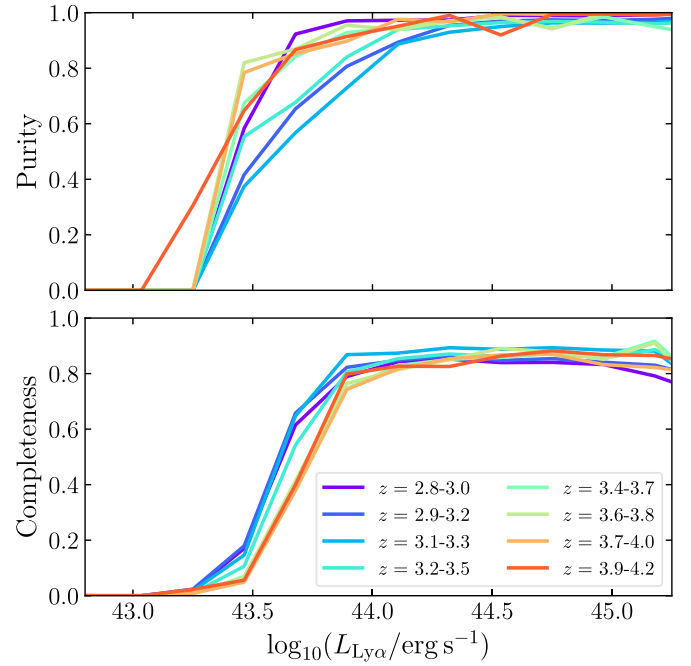
and the number count correction as

$$w^{2\text{D}} = \frac{N_{\text{tot.}}}{N_{\text{sel., true}}}, \quad (5)$$

where  $N_{\text{sel.}}$  is the number of selected objects,  $N_{\text{sel., true}}$  is the number of correctly selected objects (i.e., Ly $\alpha$  emitting quasars at the correct redshift), and  $N_{\text{tot.}}$  is the total number of objects in the mock, which would all be selected if the completeness of the method were 100%. These quantities are evaluated in each cell of the 2D grid. The estimates of  $P^{2\text{D}}$  and  $w^{2\text{D}}$  implicitly include the purity and completeness components of the NB flux excess selection and the ANN classifier. A visual inspection of our selected sample was performed only to remove candidates with observational issues which cannot be reproduced by the mocks, and thus we assume it does not affect the completeness. We also neglect the incompleteness arising from the intrinsic limitations of the reference catalogs, as these catalogs are based on significantly deeper data and are complete up to  $i \sim 24.5$  (see Sect. 2).

By marginalizing the 2D purity and completeness maps over the  $r_{\text{synth}}$  axis, we estimate the total purity and completeness of our sample in terms of  $L_{\text{Ly}\alpha}$ . The 1D purity and completeness estimations from our mocks is shown in Fig. 6. Figure C.1 is the 2D version of this figure in bins of  $r$  and  $L_{\text{Ly}\alpha}$ . The purity of the sample reaches  $\sim 90\%$  at  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \approx 44$  for all the considered redshift bins. Conversely, the completeness reaches a maximum value of  $\sim 85\%$ . Interestingly, the maximum completeness values correspond to the higher redshift bins. Fig. 6 also shows that the purity increases more rapidly with luminosity at the highest redshift bins. The reason for this is that the 7 NBs with lowest effective wavelengths are more likely to suffer from higher contamination from QSO C IV lines; the rest of the NBs probe redshifts in which the Ly $\alpha$  line is observed in combination with C IV, allowing

<sup>2</sup> Nonetheless, this is a reasonably good approximation for the PAUS NBs, and its effect should be minimal.



**Fig. 6.** One-dimensional purity and completeness of our sample as a function of measured  $L_{\text{Ly}\alpha}$ . We show the values in the bins of redshift used to estimate the Ly $\alpha$  and UV LFs (see Sect. 3.4). These curves were estimated by applying the selection method to the mocks, in bins of measured Ly $\alpha$  luminosity. The purity of the sample increases with  $L_{\text{Ly}\alpha}$ , reaching values of  $\sim 1$  for the high end at all redshifts. The completeness, on the other hand, increases monotonically with  $L_{\text{Ly}\alpha}$ , reaching maximum values of  $\approx 85\%$  at  $L_{\text{Ly}\alpha} \approx 10^{44}$  erg s $^{-1}$  at all redshifts.

our method to identify the redshift correctly. This effect reflects in Fig. 5 (top panel). This figure also illustrates that the purity of the subsample with spectroscopic counterparts is very close to unity for a large fraction of our  $z_{\text{phot}}$  range, especially for  $z_{\text{phot}} > 3.25$ , due to the decrease of line confusions.

### 3.4. Computation of the Ly $\alpha$ luminosity function

In order to compute the Ly $\alpha$  LF we follow the same scheme detailed in TT23, adapting those methods to the specific case of this work. In order to obtain the Ly $\alpha$  LF, we need to characterize the purity and completeness of our LAE candidate sample, and apply the adequate corrections. Each candidate  $k$  is assigned a weight  $w = w^{2\text{D}}$ , being  $w^{2\text{D}}$  the number count correction from Eq. (5). On the other hand, we introduce the parameter  $p_k$  whose value is zero if the  $k$ -th candidate is rejected, and unity if it is accepted to be included in the LF computation. Each candidate is accepted for the LF computation with a probability equal to the value of the purity  $P_k^{2\text{D}}$  (Eq. (4)).

The Ly $\alpha$  LF for each considered NB set is computed an arbitrarily large number of times ( $n = 1000$ ) to achieve significant statistics. The  $i$ -th determination is computed as

$$\Phi_i[\log_{10} L_{\text{Ly}\alpha}] = \frac{\sum_k p_k \cdot w_k}{V \cdot \Delta(\log_{10} L_{\text{Ly}\alpha})}, \quad (6)$$

where the index  $k$  runs through the LAE candidates within the redshift interval defined by the given NB set, in bins of Ly $\alpha$  luminosity;  $V$  is the maximum volume our survey can probe (Schmidt 1968), determined by the covered interval of wavelengths by the

NBs and the survey area<sup>3</sup>. On each iteration, the Ly $\alpha$  luminosity and  $r$  flux is randomly perturbed using Gaussian uncertainties, and the parameters  $p_k$  and  $w_k$  are recomputed accordingly. In addition, a multiplicative factor  $(1 - 0.08)^{-1}$  was applied to the LF in order to account the incompleteness due to the removal of sources with neighbours closer than 3'' from the parent sample (see Sect. 3.2.1).

In order to include the effect of cosmic variance and field-to-field variation effects, the LF is computed for bootstrapped sub-regions of the wide PAUS fields. We define tiles of equal area ( $\sim 0.12 \text{ deg}^2$ ) for each of the three PAUS wide fields. These tiles are obtained by generating random points, from an uniform distribution in spherical coordinates, within the PAUS masks and defining regions with equal number of random points inside. Finally, the LF is obtained as the median of all the  $\Phi_i$  determinations obtained from the bootstrapped regions. The final errors on our LF are computed as the 16th and 84th percentiles of the same  $\Phi_i$  distribution. Finally, the UV LF can be computed in a similar way, using Eq. (6) substituting  $\log_{10} L_{\text{Ly}\alpha}$  with the absolute UV magnitude computed at rest-frame 1450 Å,  $M_{\text{UV}}$ .

## 4. Results

In this section, we first present the composite QSO spectrum obtained from the stack of NB fluxes of the objects in our sample (Sect. 4.1). Next, we present the Ly $\alpha$  and UV luminosity functions estimated from our data (Sects. 4.2 and 4.3).

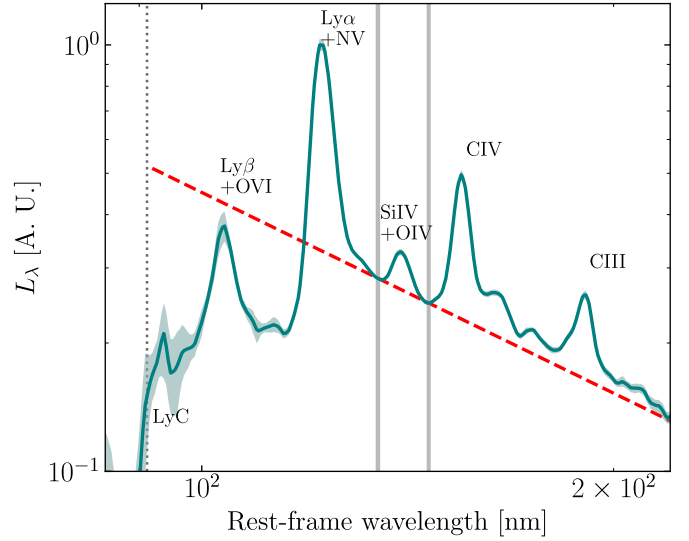
### 4.1. QSO composite spectrum

We stacked the NB SEDs of our golden sample (see Sect. 3.2.4) in order to obtain a photometric composite QSO spectrum. For this task, we use the code `stonp`<sup>4</sup>, developed to stack multi-band photospectra in a common rest-frame wavelength grid. Before stacking, each spectrum is normalized by dividing by the  $r_{\text{synth}}$  band (see Sect. 3.2.3). The composite spectrum is shown in Fig. 7, and reproduces the expected features of a QSO spectrum: power-law continuum with a break caused by the Ly $\alpha$  forest ( $\lambda_0 < 121.56 \text{ nm}$ ), and total absorption past the Lyman limit ( $\lambda_0 < 91.2 \text{ nm}$ ), and the typical Ly $\beta$ +OVI, Ly $\alpha$ +NV, SiIV+OIV, CIV, and CIII broad emission lines.

We fitted a power law to the QSO UV continuum using the rest-frame wavelength windows  $134 < (\lambda_0/\text{nm}) < 135$  and  $146 < (\lambda_0/\text{nm}) < 147$  (Fig. 7). The choice of these windows is motivated by the goal to capture the power-law shape of the spectrum in rest-frame wavelength intervals that do not contain emission lines nor any prominent noncontinuum emission, such as the Fe II emission complex (see, e.g., Vanden Berk et al. 2001). The spectral index (i.e.,  $1 + \beta_{\text{UV}}$ , being  $\beta_{\text{UV}}$  the continuum power-law slope) of QSOs is known to vary with the choice of the rest-frame wavelength window, as its measurement is susceptible to the effects of broad-emission features (e.g., thermal emission from the accretion disk, Fe II emission series, etc.; see e.g., Shull et al. 2012). For this fit, which serves as reference, we used the stack of every QSO in the full redshift range in order to maximize the S/N. We obtain a best-fitting UV slope of  $\beta_{\text{UV}} = -1.55 \pm 0.02$  for the global stack ( $f^{\lambda} \propto \lambda^{\beta_{\text{UV}}}$ ). This result is in line with the slope measured by Vanden Berk et al. (2001) for a SDSS composite spectrum using a fitting window of  $135 < (\lambda_0/\text{nm}) < 136.5$  ( $\beta_{\text{UV}} = -1.56$ ). More recently, Lusso et al.

<sup>3</sup> The covered volume  $V$  is computed by integrating the Astropy function `differential_comoving_volume` over the redshift interval covered by each NB, and multiplying by the effective survey area.

<sup>4</sup> <https://github.com/PAU-survey/stonp>



**Fig. 7.** Stacked composite spectrum of all QSOs in our golden sample. We show the mean composite spectrum (teal solid line) and bootstrap  $1\sigma$  interval (shaded region). The shaded gray regions mark the rest-frame wavelength intervals used to fit a power-law continuum. We represent the fitted power law as a dashed red line, whose slope is  $\beta_{\text{UV}} = -1.55 \pm 0.02$ . We show the labels of the most important QSO emission lines and the LyC limit (91.2 nm).

(2015) measured a UV slope of  $\beta_{\text{UV}} = -1.61 \pm 0.01$ , using a different choice for the wavelength windows (109.5–111.0, 113.5–115.0, 145.0–147.0, 197.5–201.0, and 215.0–220.0 nm).

### 4.2. The Ly $\alpha$ luminosity function at $2.7 \lesssim z \lesssim 5.3$

Figure 8 shows the Ly $\alpha$  LF in the nine overlapping intervals of redshift as defined in Table 1. We discuss the evolution of our LF in Sect. 5.1 below. The highest redshift bin in Fig. 8 is computed from the high- $z$  sample (see Sect. 3.2.5), and assuming purity and completeness equal to 1. The reason for this is that our 2D purity and number-count correction estimations cannot be accurately computed for  $z \gtrsim 4.5$ , where the SDSS DR16 sample begins to be scarce. As discussed in Appendix C of TT23, purity inaccuracies do not significantly affect the estimation of the LF, and completeness smaller than 1 induces a positive correcting factor in the LF. Hence, the results of this highest- $z$  bin should be taken as lower limits to the Ly $\alpha$  LF. We also note that, due to the limited number of candidates for such high redshifts, this last bin is considerably wider than the rest, covering  $z \approx 4.26\text{--}5.26$ .

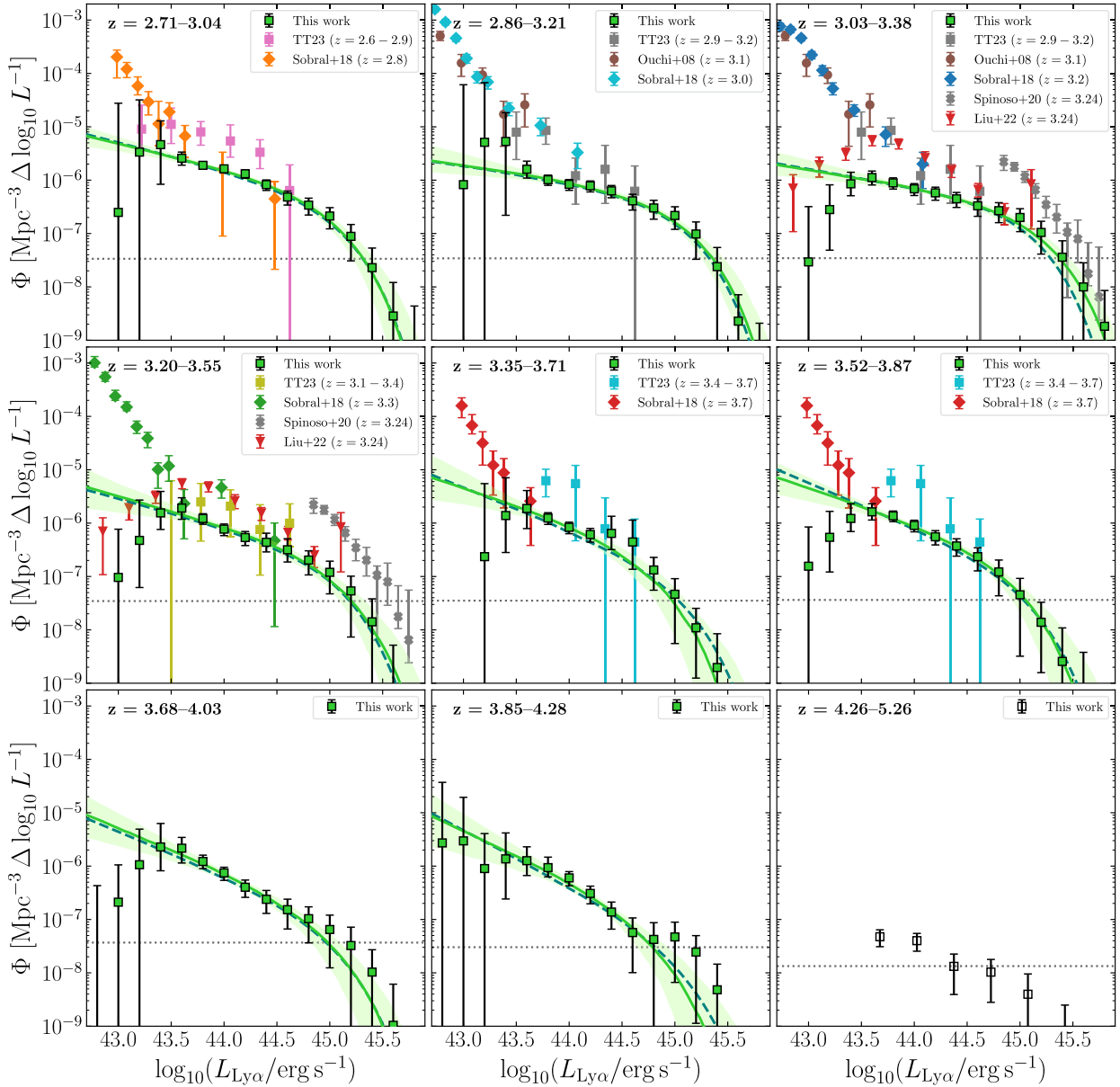
#### 4.2.1. Schechter fit to our data

We fitted a Schechter function (Schechter 1976) to our data, of the form

$$\Phi[L]d(\log_{10} L) = \log 10 \cdot \Phi^* \cdot \left(\frac{L}{L^*}\right)^{\alpha+1} \cdot e^{-L/L^*} d(\log_{10} L). \quad (7)$$

We performed the fitting using the Monte-Carlo Markov Chain (MCMC) technique, aided by the Python package `autoemcee`<sup>5</sup>. We fitted our Ly $\alpha$  LF in the full measured range for consistency. In all panels of Fig. 8, a deviation from the characteristic power-law shape of the Schechter function is evident at the faint end of the LFs. This effect is likely due to inaccuracies in the completeness and purity models at such low luminosity intervals.

<sup>5</sup> <https://github.com/JohannesBuchner/autoemcee>



**Fig. 8.**  $\text{Ly}\alpha$  luminosity function in nine intervals of redshift defined in Table 1. We compare with  $\text{Ly}\alpha$  LFs in the literature at similar redshifts. The green solid line and blue shaded area show the best Schechter fit with free parameters and the  $1\sigma$  confidence interval, respectively. The dashed blue line shows the Schechter best fit for fixed  $L^* = 45 \text{ erg s}^{-1}$ . The  $\text{Ly}\alpha$  LF of the highest-redshift bin shown in this figure was computed assuming purity and completeness equal to 1, and therefore it should be regarded as a lower limit (see Sect. 4.2). The horizontal gray dotted line marks the 1 object per bin level. Data points below this limit signal LF bins that contain less than one object after we average out the 1000 LF determinations used to compute our final LF.

However, our method effectively accounts for these inaccuracies by generating large uncertainties, which reduce the influence of these bins on the fitted curve.

The Schechter fit was performed using a normal log-likelihood and wide flat priors for all three free parameters:  $\log_{10}(\Phi^*/\text{Mpc}^{-3}) \in [-9, -3]$ ,  $\log_{10}(L^*/\text{erg s}^{-1}) \in [43, 46]$  and  $\alpha \in [-3, -1]$ . The best-fitting parameters are summarized in Table D.1. The best-fitting value for the normalization  $\Phi^*$  marginally shows anti-correlation with redshift. The effect of this evolution can be easily seen as the overall decline towards higher redshift of the  $\text{Ly}\alpha$  LF in Fig. 8. The faint-end power-law slope  $\alpha$  also shows steepening values with redshift, while the characteristic luminosity  $L^*$  has no clear evolutionary trend in the probed redshift range.

We performed further Schechter fits to our  $\text{Ly}\alpha$  LFs fixing  $\log_{10} L^* = 45 \text{ erg s}^{-1}$ . Fixing this parameter removes the possible correlations with  $\Phi^*$  ( $\Phi^*$  defines the value of the LF at  $L = L^*$ , see Eq. (7)). The evolution of the parameter  $\Phi^*$  becomes more evident, and the faint-end slope shows a very similar steepening trend with redshift. We discuss the fitted parameters more in detail in Sect. 5.1.

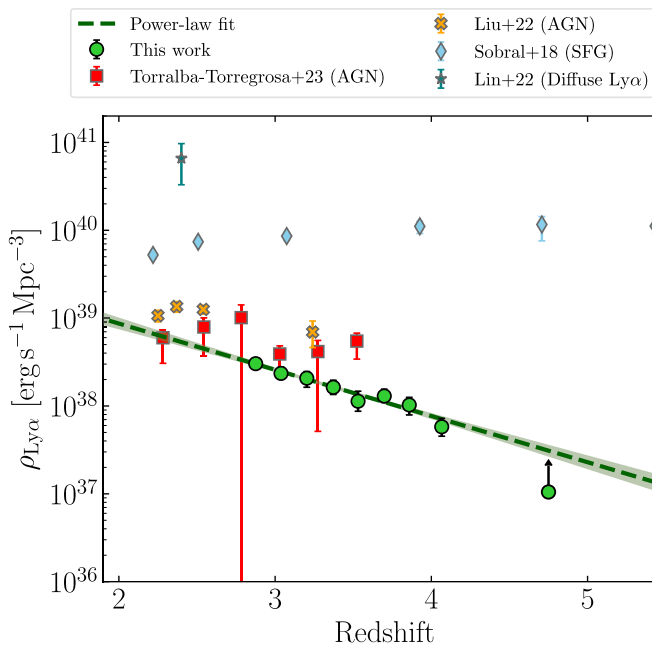
#### 4.2.2. Integrated $\text{Ly}\alpha$ luminosity function

In Fig. 9, we show the results of integrating the best fit of the  $\text{Ly}\alpha$  luminosity function in the range  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) > 43.5$ , for the redshift intervals shown in Fig. 8. The resulting total volume density of  $\text{Ly}\alpha$  emission  $\rho_{\text{Ly}\alpha}$  shows a clear

**Table 1.** Intervals of NBs used for the Ly $\alpha$  and UV LF computation.

NB <sub>min</sub>	NB <sub>max</sub>	z <sub>min</sub>	z <sub>max</sub>	Volume [10 <sup>6</sup> Mpc <sup>3</sup> ]
0	3	2.71	3.04	4.167
2	5	2.86	3.21	4.089
4	7	3.03	3.38	4.090
6	9	3.20	3.55	4.044
8	11	3.35	3.71	3.975
10	13	3.52	3.87	3.855
12	15	3.68	4.03	3.799
14	18	3.85	4.28	4.634
19	30	4.26	5.26	10.481

**Notes.** The NB indices shown in the first two columns refer to the NB IDs defined in Table A.1.



**Fig. 9.** Ly $\alpha$  luminosity density per unit comoving volume. These points were obtained by integrating the Ly $\alpha$  LF for  $L_{\text{Ly}\alpha} > 10^{43.5} \text{ erg s}^{-1}$ . Our results show a declining trend with increasing redshift, compatible with the results presented in TT23 and with the integrated Ly $\alpha$  LF from Liu et al. (2022). We also compare with the integrated LF of the SFG population in Sobral et al. (2018), which covers  $L_{\text{Ly}\alpha} \lesssim 10^{43.5} \text{ erg s}^{-1}$ , showing that the integrated AGN Ly $\alpha$  luminosity is increasingly smaller than the galaxy contribution at higher redshift. We also compare with the diffuse Ly $\alpha$  emission measurement of Lin et al. (2022).

declining trend towards increasing redshifts. We fit this trend to  $\log_{10}(\rho_{\text{Ly}\alpha}/\text{erg s}^{-1} \text{ Mpc}^{-3}) = b + a \cdot z$ , obtaining the parameters  $a = -0.53 \pm 0.05$  and  $b = 39.90 \pm 0.16$ . For this fit, we excluded the highest redshift bin (empty circle in Fig. 9), however this lower-limit is in agreement with the extrapolation of the fitted evolution of  $\rho_{\text{Ly}\alpha}$ .

Our results are compatible with the measurements obtained by integrating the best-fits of the data presented in TT23, for most of the redshift bins. The surveyed area employed in TT23 was significantly smaller (by a factor  $\sim 30$ ), and the discrepancies can be due to cosmic variance. The extrapolation of our  $\rho_{\text{Ly}\alpha}$  best-fit is considerably lower than the result of integrating the double power-law fits in Liu et al. (2022), in their  $z = 2.25, 2.37$  and  $2.54$  bins, in the same luminosity range. These differences are also reflected in the discrepancies with their faint-end

of the Ly $\alpha$  LF that can be seen in Fig. 8. In their work, they used blind spectroscopy to select a sample of much fainter AGNs, which also include narrow-line AGNs and sources with significant contribution from star formation in the host galaxy, while our sample is mainly composed of luminous broad-line quasars.

We also compare the presented measurements of  $\rho_{\text{Ly}\alpha}$  for the AGN Ly $\alpha$  LF with the measurements of Sobral et al. (2018), for the faint end of the Ly $\alpha$  LF, populated by star-forming galaxies. The value of  $\rho_{\text{Ly}\alpha}$  for this population increases with redshift, showing the opposite trend to the bright Ly $\alpha$  emitting AGN population. Our results show that for  $z \approx 4$ , the total emissivity of AGN Ly $\alpha$  is subdominant by a factor of  $\approx 200$ , with respect to the star-forming LAEs. Moreover, we also display the estimated  $\rho_{\text{Ly}\alpha}$  from unresolved emitters and diffuse IGM emission from Lin et al. (2022); this value is expected to be significantly larger than Ly $\alpha$  luminosity from resolved emitters (Byrohl & Nelson 2023), but can be explained by integration of faint LAE LFs (e.g., Drake et al. 2017; Sobral et al. 2018) to an arbitrarily faint end (Renard et al. 2024).

### 4.3. UV luminosity function

One of the advantages of our multi-NB data is the possibility to measure SEDs for all the objects in the field without preselection, in an extensive range of wavelengths. We measure the UV absolute magnitude ( $M_{\text{UV}}$ ) for every object in our selected sample, defined as the luminosity at rest-frame wavelength 145 nm (e.g., Glikman et al. 2010). The UV magnitude is thus computed as

$$M_{\text{UV}} = m_{\text{NB}} - \text{DM}(z) - K, \quad (8)$$

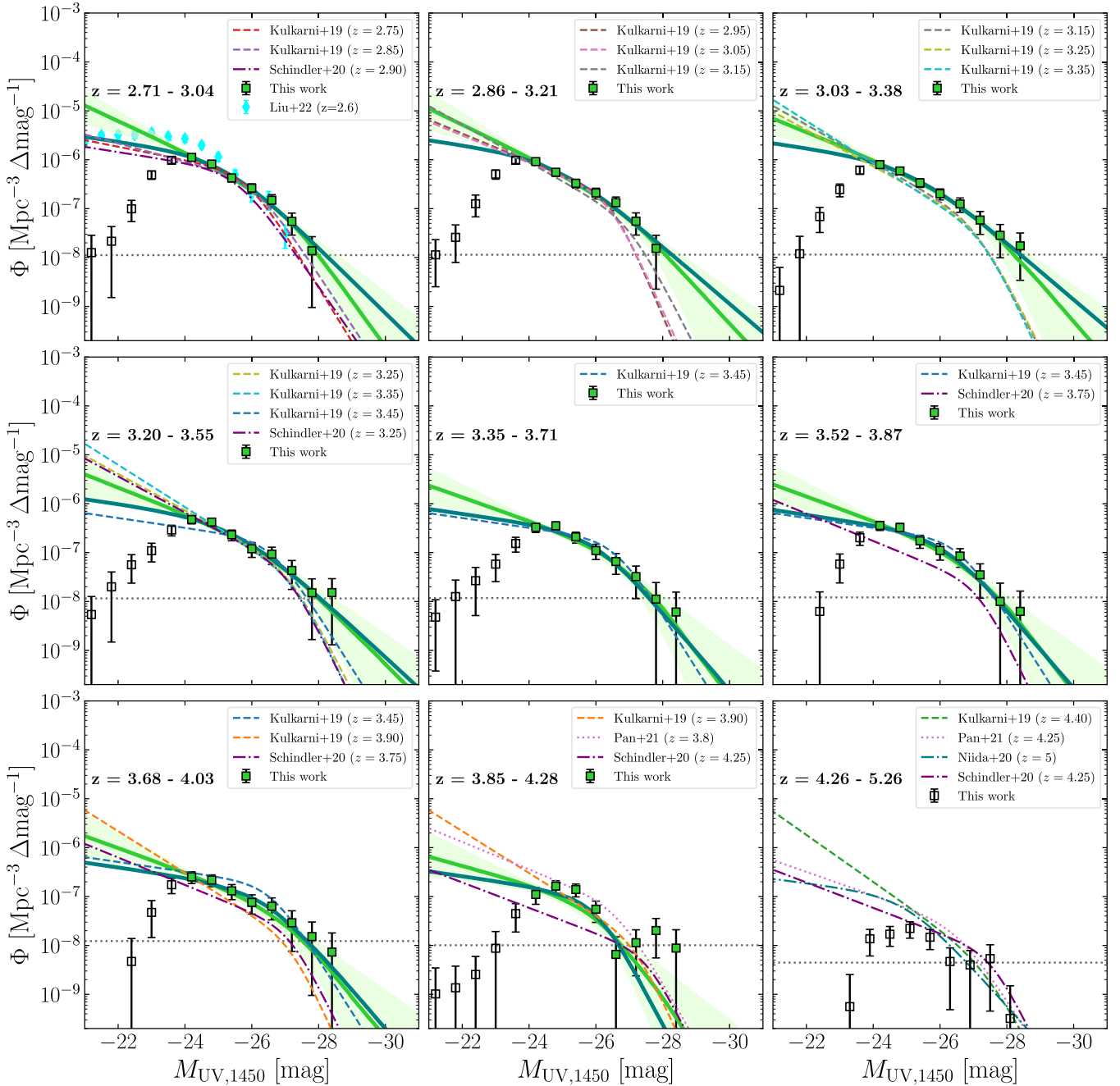
where  $m_{\text{NB}}$  is the apparent magnitude of the NB corresponding to the rest-frame wavelength 145 nm of each object,  $\text{DM}(z)$  the redshift dependent distance modulus according to our cosmology, and  $K$  the  $K$ -correction defined as

$$K = -2.5 \log_{10}(1 + z)^{\alpha+1}, \quad (9)$$

where  $\alpha$  is the spectral index (i.e.,  $\alpha = -2 - \beta_{\text{UV}}$ ; see e.g., Wisotzki 2000). For the computation of the  $K$ -correction we adopt the UV slope measured in our NB composite spectrum ( $\beta_{\text{UV}} = -1.55$ ,  $\alpha = -0.45$ ; see Sect. 4.1). We note that the continuous coverage of the PAUS filters allow us to photometrically measure  $M_{\text{UV}}$  in the adequate window without having to apply bandpass corrections.

We obtain the UV LFs for our sample following a very similar procedure to that used for obtaining the Ly $\alpha$  LF, described in Sect. 3.4. The median purity and number count corrections based on  $r_{\text{synth}} - L_{\text{Ly}\alpha}$  are applied, because the selection of the sources which make part of the UV LF is still based on the same selection function as the Ly $\alpha$  LF. Hence, our results presented in Fig. 10, correspond to the UV LFs of our Ly $\alpha$ -selected AGN sample. We compare our results with the best-fitting curves in the literature for similar redshift intervals (Kulkarni et al. 2019; Niida et al. 2020; Schindler et al. 2019; Pan et al. 2022).

The last redshift bin of our UV LF ( $4.26 < z < 5.26$ ) is computed in the same way as for the Ly $\alpha$  LF (see Sect. 4.2), using visually selected objects from a preselection by our method (see Sect. 3.2.5). As for the case of the Ly $\alpha$  LF (see introduction of Sect. 4.2), the purity and completeness of this redshift bin is not well determined, and therefore this result may be taken as a lower limit for the UV LF. Nevertheless, the UV LF in this last redshift bin is in good agreement with  $z \sim 5$  determinations in the literature (Kulkarni et al. 2019; Niida et al. 2020; Pan et al. 2022).



**Fig. 10.** UV luminosity function of our sample selected via Ly $\alpha$  emission. We show the LF in the same redshift bins of the Ly $\alpha$  LF (Fig. 8). The filled green squares are used to fit a double power law, while the empty squares are discarded in the fit, due to low completeness (average completeness <50%). The solid green line and shaded area show the best-fit and  $1\sigma$  uncertainties to a double power-law (teal solid line shows the fit with fixed  $\beta$ ). We compare with double power-law fits in the literature, for the redshifts that approximately fall in our bins. The UV LF of the highest redshift bin shown in this figure was computed assuming completeness equal to 1, and therefore it should be regarded as a lower limit (see also Fig. 8 and Sect. 4.2). We mark the 1 object per bin limit as a horizontal dotted line. Data points below this limit signal LF bins which contain less than 1 object after we take the average of the 1000 LF determinations used to compute our final LF.

#### 4.3.1. Double power law fit

The faint end of our UV LF ( $M_{UV} \gtrsim -24$ ) is affected by the poor purity and completeness (see Fig. 6). For  $M_{UV} \lesssim -24$ , the UV LF is typically described by a double-power law (DPL):

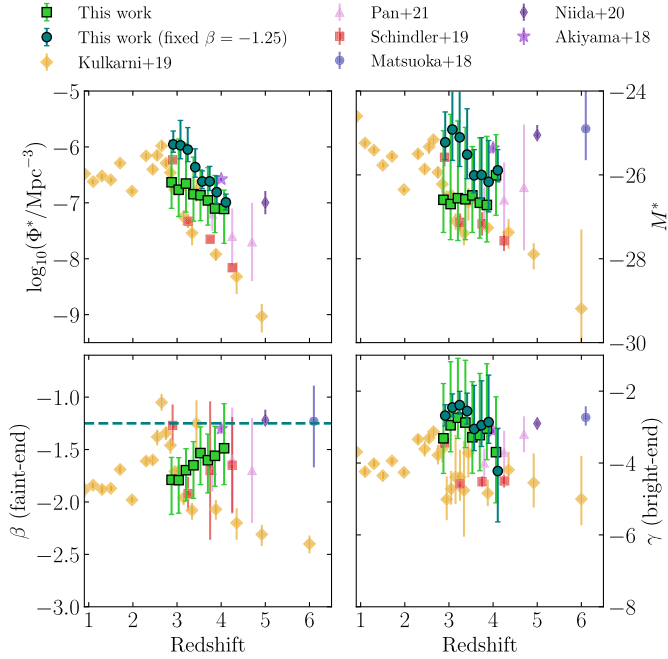
$$\Phi[M] = \frac{\Phi^*}{10^{-0.4(M^*-M)(1-\beta)} + 10^{-0.4(M^*-M)(1-\gamma)}}. \quad (10)$$

We fit our UV LF using Eq. (10), using flat priors:  $\log_{10}(\Phi^*/\text{Mpc}^{-1}) \in [-10, -5]$ ,  $M^* \in [-28, -22]$ ,  $\beta \in [-3, -1]$ ,

$\gamma \in [-6, -1]$ . We obtain the best-fit parameters shown in Fig. 11. The fitted parameters are discussed in Sect. 5.2 below.

#### 4.3.2. Integrated UV emissivity

The integral of the UV LF directly yields the total UV emissivity of the AGN population per unit volume. We integrate our best fits to obtain  $\epsilon_{1450}$ , the emissivity at 145 nm (i.e., the wavelength where  $M_{UV}$  is measured), by extrapolating from  $M_{UV} = -21$  to an arbitrarily luminous magnitude. This allows



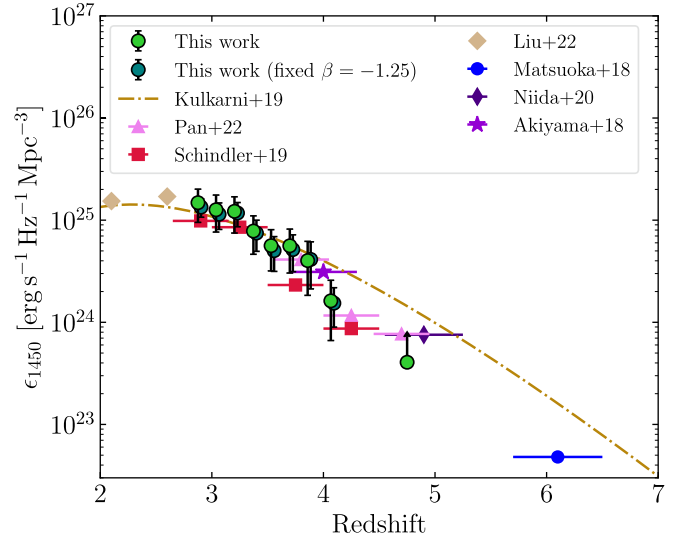
**Fig. 11.** Best-fit double power-law parameters for the UV LF in bins of redshift. We show the results obtained by leaving all four parameters free (green squares) and fixing the faint-end slope ( $\beta = -1.25$ , shown as a horizontal dashed blue line; blue circles). We compare our results with those of Kulkarni et al. (2019). Our results show a clear decrease in the value of the normalization parameter  $\Phi^*$  with redshift, in line with the fitted Schechter parameter for the Ly $\alpha$  LF. The faint-end slope  $\beta$  also shows hints of a declining evolution, while the bright-end slope  $\gamma$  is poorly constrained by our data.

for a direct comparison with Fig. 8 in Kulkarni et al. (2019)<sup>6</sup>, as we show in our Fig. 12. Our Results are compatible with the fitted trend in Kulkarni et al. (2019) as well as with the integrated UV LFs of several past works in the literature. We computed  $\epsilon_{1450}$  in the same range for our best fits with fixed faint end slope ( $\beta = -1.25$ ), obtaining similar values. Therefore, a small change in the faint-end slope has little impact on the total emissivity of intermediate-luminosity quasars.

#### 4.4. Mean Lyman- $\alpha$ forest transmission

In Sect. 4.1 we stacked our sample in order to obtain a composite QSO spectrum in the full redshift range ( $z = 2.7\text{--}5.3$ ). Here we repeat the same procedure in bins of redshift. In Fig. 13 we show the stacked spectra in 8 bins of redshift. At wavelengths redward Ly $\alpha$ , our QSO spectra are self-similar at all redshifts – the average UV spectrum of quasars is known to not significantly vary across cosmic time (e.g., Onorato et al. 2024). The EW of the most luminous emission lines (Ly $\alpha$  and C IV) is notably scattered, however this is probably due to fainter sources having higher S/N at lower redshifts, dominating the sample. The uncertainties in the systemic redshift can induce biases in the emission-line EWs of the stacked spectra. In addition, the spectral resolution of the NB pseudo-spectra increases with redshift, given that the rest-frame wavelength interval that is covered by a single NB is proportional to  $\propto(1+z)$ . For this reason, the shape of the Ly $\alpha$  line might also vary in the composite spectrum, once

<sup>6</sup> In Kulkarni et al. (2019), the 91.2 nm emissivity is shown instead. This value is obtained extrapolating an assumed power-law continuum. We choose to directly show the 145 nm emissivity, rescaling the values shown in their Fig. 8.



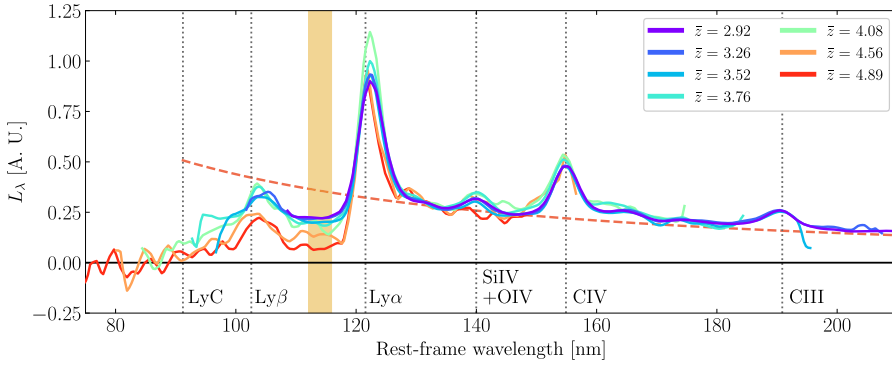
**Fig. 12.** UV emissivity at rest-frame wavelength 1450 Å. We show the emissivity obtained from the integral of the double power-law fit with four free parameters (green circles) and fixed faint end slope  $-1.25$  (blue circles) – the latter are slightly shifted in the horizontal axis for clarity. The total emissivity at 145 nm rapidly declines with redshift. This quantity is directly related to the ionizing photon emissivity (i.e.,  $\lambda_0 < 91.2$  nm), given that the AGN continuum is well approximated by a power-law. We compare our results with the integral of other UV LFs in the literature for  $z > 2$  (Matsuoka et al. 2018; Akiyama et al. 2018; Schindler et al. 2019; Kulkarni et al. 2019; Niida et al. 2020; Pan et al. 2022; Liu et al. 2022).

the latter is brought to the rest-frame (e.g., the Ly $\alpha$  lines appear to be narrower in the two highest redshift bins).

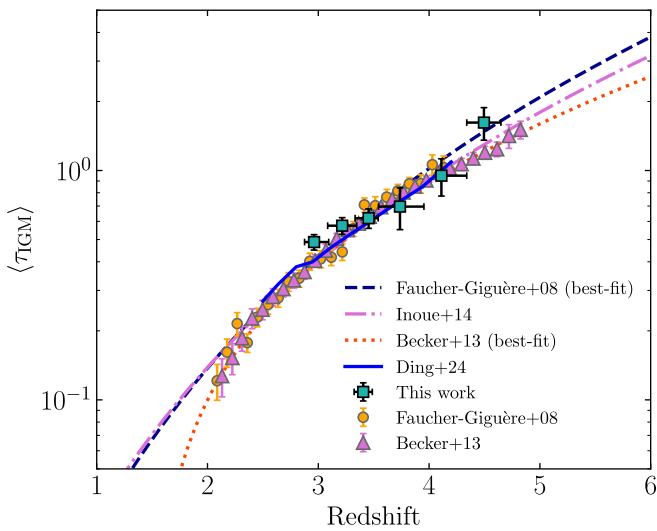
On the other hand, the continuum in the rest-frame wavelength range within  $91.2 < (\lambda_0/\text{nm}) < 121.57$  shows an evolution with redshift consistent with increasing mean IGM absorption by the Ly $\alpha$  forest (+Ly $\beta$  forest for  $\lambda_0 < 102.5$  nm). We average the Ly $\alpha$  forest transmitted flux of the composite spectra in each redshift bin, in the range  $113 < (\lambda_0/\text{nm}) < 115$  (wavelength interval in which no prominent emission lines are expected; see Fig. 13), and divide it by the extrapolated power-law flux fitted above, in order to obtain the mean IGM absorption due to the Ly $\alpha$  forest. We show the mean IGM optical depth ( $\tau_{\text{IGM}}$ ) in Fig. 14, which relates to the IGM transmission as  $T_{\text{IGM}} = e^{-\tau_{\text{IGM}}}$ . For this analysis, we only consider redshifts where the observed window used to estimate the transmitted flux is covered at least by the second NB in our set (NB465), to avoid border effects of the stack. Our measurements are in line with previous determinations in the literature (Faucher-Giguère et al. 2008; Becker et al. 2013; Inoue et al. 2014; Ding et al. 2024) in the range  $z = 3\text{--}5$ . We note that the best fit in Faucher-Giguère et al. (2008) was used in our selection method in order to correct the  $\text{EW}_{\text{Ly}\alpha}$  measurement, and Fig. 14 shows that our selected sample retrieves the same result.

## 5. Discussion

In this section, we compare our estimated redshift-dependent Ly $\alpha$  and UV LFs with past measurements in the literature, and discuss the evolution of the QSO number densities as a function of cosmic time (Sects. 5.1, 5.2). We comment on the efficiency of our selection method to retrieve a complete and pure sample of Ly $\alpha$ -selected QSOs (Sect. 5.3), and cross-match our sample with spectroscopic surveys (Sect. 5.4).



**Fig. 13.** Composite QSO spectra in different bins of redshift. The composite spectra were obtained stacking the objects in our visually selected sample using the `stomp` code. The UV continuum slope show negligible evolution with redshift. The visible differences in the typical EW of the most prominent lines is likely due to intrinsically fainter sources dominating the lowest redshift bins, and the resolution increase in the rest-frame wavelength with redshift. The dotted vertical lines mark the rest-frame wavelengths of the most prominent QSO lines, and the Lyman limit (91.2 nm; LyC). The dashed pink line shows the best power-law fit for the UV continuum, with slope  $\beta_{UV} = -1.55 \pm 0.02$ . The shaded orange region is used to compute the average IGM optical depth (see Sect. 4.4).



**Fig. 14.** Mean IGM transmission due to the  $\text{Ly}\alpha$  forest. These estimations were obtained from the stacked composite spectra shown in Fig. 13. As comparison, we show the estimations in Becker et al. (2013), Inoue et al. (2014), Ding et al. (2024) and Faucher-Giguère et al. (2008), the latter employed as continuum correction in our selection method (see Sect. 3.2).

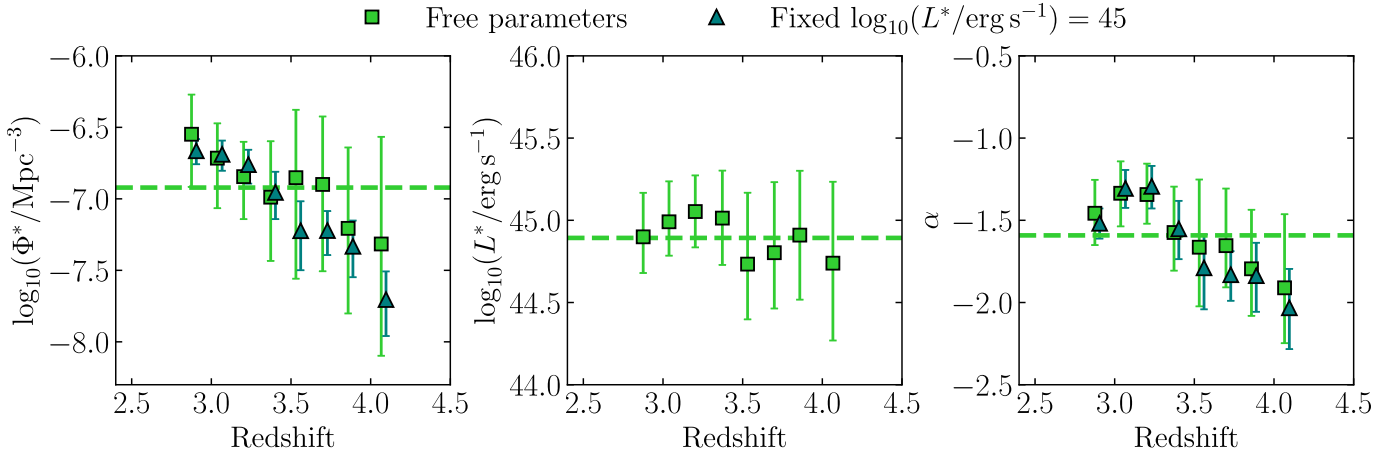
### 5.1. $\text{Ly}\alpha$ LF: Evolution and comparison with literature

As described in Sect. 4.2.1, we used a Schechter function to fit our  $\text{Ly}\alpha$  LFs (e.g., Spinoso et al. 2020, TT23). Contrary to previous works, our data probing the faint end ( $L_{\text{Ly}\alpha} < L^*$ ) enables to sample three Schechter parameters with good significance. Nevertheless, the correlations between these parameters are strong, and thus we choose to fix the turnover luminosity parameter  $L^*$ , in order to have a better understanding of the evolution of  $\alpha$  and  $\Phi^*$ . As shown in Fig. 15, the normalization  $\Phi^*$  decreases with redshift – this being more evident for a fixed  $L^*$ . This result particularly aligns with the evolution of the total SMBH mass function predicted by semi-analytical models (e.g., Fanidakis et al. 2012; Izquierdo-Villalba et al. 2020). In addition, in our results the faint-end slope  $\alpha$  evolves toward steeper values with redshift. This trend suggests that, at lower redshift, faint AGNs tend to be proportionally less abundant than the brighter analogs ( $L_{\text{Ly}\alpha} > L^*$ ). Interestingly, our results about the evolution of  $\Phi^*$  and  $\alpha$  appear to be robust against the choice of fixing  $L^*$ , suggesting that the trends are likely not driven by a shift in the overall luminosity of the AGN population, but instead to be due an evolving balance between faint and bright AGN.

In general, our results are consistent with previous estimations of the AGN  $\text{Ly}\alpha$  LF in the literature. In particular, the results of this work are in agreement with TT23 within the large intervals of confidence, demonstrating the consistency of our method for different datasets. The sky area in TT23 is considerably smaller ( $1.14 \text{ deg}^2$ ) than the combined  $\sim 35 \text{ deg}^2$  of the PAUS wide fields, and the depth of both surveys is similar. For this reason, in this work we are able to measure the LF in a range of higher  $\text{Ly}\alpha$  luminosities; pushing toward the bright end and higher redshifts. The main limiting factor at the LF bright-end is the rapid drop of bright AGNs at increasing redshifts (e.g., Kulkarni et al. 2019; Shen et al. 2020). The discrepancies between this work and some redshift bins in TT23 can be explained by the cosmic variance due to the small area, and higher contamination in the TT23 sample product of a less refined selection method.

Our estimate of the binned LF differs from that of Sobral et al. (2018) and Liu et al. (2022) in the intermediate luminosity regime ( $\log_{10}(L_{\text{Ly}\alpha}) \sim 43.5\text{--}44 \text{ erg s}^{-1}$ ). Sobral et al. (2018) uses a sample selected in deep medium and narrow bands (magnitude  $\sim 23.5$  and  $\sim 25.5\text{--}26$  at  $5\sigma$ , respectively) within a smaller area of  $\sim 2 \text{ deg}^2$ . Their sample contains mainly SFGs, with a small fraction of sources that present radio or X-rays counterparts ( $\sim 1.2\text{--}4.4\%$ ), suggestive of AGN activity. Our best-fits for faint end slope (ranging from  $\alpha = -1.45$  to  $-1.91$ ; mean  $-1.59$ ) are in line with the slopes yielded by the results in Sobral et al. (2018) which employed a combination of a Schechter ( $\alpha = -1.7^{+0.3}_{-0.2}$ ) and a power-law ( $A-1 = -1.75 \pm 0.17$ ), where the latter only fitted the  $\text{Ly}\alpha$  LF of the objects showing AGN radio or X-ray signatures. These fits were obtained from a sample that mainly contains objects in the  $\text{Ly}\alpha$  luminosity range of  $42.5 \leq \log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \leq 43.5$  (for comparison, our luminosity range starts at  $\sim 43.5 \text{ erg s}^{-1}$ ).

In Liu et al. (2022) the  $\text{Ly}\alpha$  LF is estimated from a sample of objects selected through blind spectroscopy in HETDEX (Gebhardt et al. 2021). Our  $\text{Ly}\alpha$  LF at  $z \sim 3.24$  is significantly lower than the one of Liu et al. (2022). Their sample contains fainter (up to  $r \sim 26$ ) objects with typically narrower lines, which might contain significant contribution from star-formation activity and Type-II AGN (e.g., Hainline et al. 2011). However, the selection method employed by Liu et al. (2022) relies on the detection of two emission lines inside a limited spectral range, which limits the completeness in specific intervals of redshift where two or more QSO lines are not visible. In addition, their selection method is less efficient in detecting very broad lines, and thus their sample suffers from incompleteness in the brightest  $\text{Ly}\alpha$  luminosity bins. Our results capture better



**Fig. 15.** Best-fit Schechter parameters for the LFs in the nine intervals of redshift. The horizontal dashed lines represent the average best-fit Schechter parameters across all redshifts. When leaving all three parameters free (green squares), the normalization  $\Phi^*$  and the characteristic luminosity  $L^*$  show a slight decline within the intervals of confidence, whilst the faint-end slope  $\alpha$  shows no clear trend. We also show the fitted parameters when fixing  $\alpha$  and  $L^*$  to the average values marked with the green dashed lines (blue triangles). The values of the latter are slightly shifted in the horizontal axis for visual clarity.

the exponential Schechter decay of the Ly $\alpha$  LF at the bright end ( $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) > 44.5$ ).

Finally, our bright-end of Ly $\alpha$  LF strongly disagrees with that of Spinoso et al. (2020) in the  $z = 3.24$  bin (their highest probed redshift). In Spinoso et al. (2020) the LF is estimated from a sample of NB-selected QSOs in the  $\sim 1000 \text{ deg}^2$  of the J-PLUS survey (Cenarro et al. 2019) data-release one. The discrepancies with our estimation of the Ly $\alpha$  LF at  $z = 3.24$  can be explained by the fact that, as stated in Spinoso et al. (2020), their  $z \sim 3.24$  QSO sample is highly affected by mis-classification of C IV or C III QSO lines (at  $z < 3$ ) as Ly $\alpha$  at  $z \sim 3.24$ . In addition, the J-PLUS data used in their work had a limiting depth of  $r \lesssim 21.5$ , which made difficult a reliable identification of  $z \sim 3.24$  objects. Finally, as reported in their work, their purity computation was based on a cross-match with SDSS QSO catalogs, which become increasingly incomplete at  $z > 3$ , likely leading to an overestimated purity for the highest-redshift bin they analyzed.

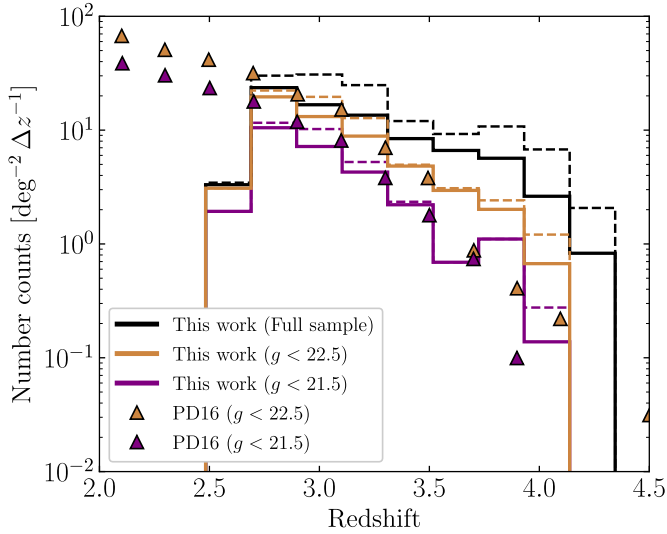
## 5.2. UV LF: Evolution and comparison with literature

The UV LF of quasars has been intensively studied in the literature, mostly focusing on spectroscopically selected samples (e.g., Matsuoka et al. 2018; Akiyama et al. 2018; Kulkarni et al. 2019; Schindler et al. 2019; Niida et al. 2020; Pan et al. 2022). The UV LF clearly evolves across cosmic time, and several models have been proposed to explain this evolution: from “pure luminosity evolution” (PLE) models, where the change in the LF is mainly driven by a shift in the luminosity distribution; to “luminosity evolution and density evolution” models where both the luminosity distribution and global number density of AGNs evolve (see e.g., Boyle et al. 2000; Ross et al. 2013).

Our best-fitting DPL parameters reflect an evolution of the UV LF (see Fig. 11). The normalization parameter  $\Phi^*$  in general decreases with redshift, in parallel with what is found for our Ly $\alpha$  LF Schechter fits (Sect. 5.1). This decrease is in line with the evolution found by works in the literature that studied the UV LF at different redshifts, as is also made evident in Fig. 12. This trend also becomes more evident when fixing the faint-end slope. This result is consistent with a global picture in which cosmic AGN activity decreases from  $z \sim 2$  toward  $z \sim 5$  (e.g.,

Shen et al. 2020). When left free, the faint-end slope  $\beta$  shows a correlation with redshift, evolving toward flatter slopes. However, we note that the faint-end of the UV LF might not be well constrained by our data. In general, this parameter does not show consistent evolution in the literature (see Fig. 11). The other two DPL parameters ( $\gamma, M^*$ ) do not present an evident evolution. In particular the bright-end slope  $\gamma$  shows rather shallow values, as compared to the ones obtained by previous works. The excess in our UV LF is not significant given the uncertainties, and can be explained by cosmic variance. This excess cannot be explained by magnification bias, as this effect is predicted to become relevant at  $\sim 8$  (Wyithe et al. 2011). In general all the DPL parameters present significant correlations and large variance.

In Akiyama et al. (2018), Niida et al. (2020) and Matsuoka et al. (2018), the UV LF of quasars is computed using a combination of deep imaging data for faint AGNs ( $M_{\text{UV}} \gtrsim -26$ ) and SDSS samples of quasars for the bright-end ( $M_{\text{UV}} \lesssim -26$ ). These three works fitted significantly flatter faint-end slopes (mean  $\beta = -1.25$ ; see Fig. 11, with no significant evolution with redshift. On the other hand, Kulkarni et al. 2019) obtained steeper faint-end slopes, by only employing QSOs from SDSS and BOSS, which in general are more luminous than the sample of HSC deep imaging. In order to try to reconcile our results with the faint-end slope constrained by HSC, we fit a DPL model to our UV LFs using a fixed  $\beta = -1.25$ . The fitted curve is displayed in Fig. 10 as a teal solid line, and it is generally compatible with our binned LF, showing that our data does not tightly constrain the faint-end slope. The fits yield slightly fainter values for the turnover magnitude  $M^*$ , which are more in agreement with the fainter values for  $M^*$  derived from the HSC+SDSS data. By fixing  $\beta = -1.25$  we also obtain better constrained values of the bright-end slope, and more in line with the HSC+SDSS determinations. Moreover, the anticorrelation of  $\Phi^*$  with redshift increases, reflecting an evolution consistent with that of the Ly $\alpha$  LF (see Sect. 4.2.1). The fact that fixing  $\beta = -1.25$  yields fainter turnover magnitudes might mean that our UV LF only covers a specific  $M_{\text{UV}}$  range (i.e.,  $M_{\text{UV}} \sim M^*$ ), hence hindering our ability to well constrain the faint end of the power-law LF, which is instead well-sampled in the HSC data.



**Fig. 16.** Differential number counts of QSOs per unit area of sky. The dashed histograms show our complete selected sample, the solid histograms show the distribution of our visually selected golden sample. We compare with the spectroscopic sample of Palanque-Delabrouille et al. (2016, PD16), selected through QSO variability. Our number counts are significantly higher than those of PD16, especially at the highest bins of redshift.

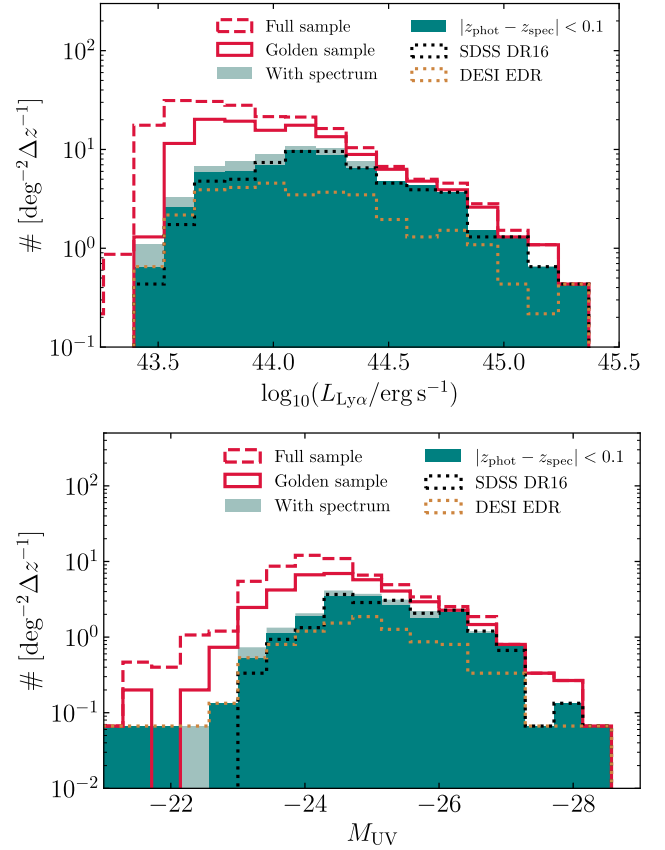
### 5.3. Efficiency of the selection method

The selection method employed in this work (see Sect. 3.2) effectively selects quasars based on strong line emission, selected mainly through NB flux excess. We do not apply any direct color cut in our selection, which removes biases that affect most spectroscopic samples that rely on photometric pre-selection on color-color diagrams (e.g., Schneider et al. 2010; Ross et al. 2013; Lyke et al. 2020). Selection methods based on QSO variability have been proved to be capable to obtain more pure and complete samples than color-based selections, however loose color cuts are often also applied in order to minimize contamination from stars (e.g., Palanque-Delabrouille et al. 2011). The problems introduced by color cuts especially affect the redshift range  $2.5 \lesssim z \lesssim 3.5$ , where most QSOs occupy the sample region as the stellar locus in optical color-color planes (Fan 1999), and the selection functions become inefficient (see e.g., Richards et al. 2006; Ross et al. 2013). Moreover, variability in quasars anti-correlates with continuum luminosity and redshift, and the variability of emission lines is found to be only  $\sim 10\%$  of the variability of the continuum (Meusinger et al. 2011). In Schindler et al. (2017) it is shown that the classical QSO selection methods employed by SDSS are prone to miss a significant part of  $z > 3$  quasars in the brightest end, mainly due to the inefficiency of the color-based selection.

In Fig. 16 we show the number counts per unit area in different  $g$ -band cuts for  $2.75 < z < 4.3$ . Our number counts are in line with those of Palanque-Delabrouille et al. (2016) for  $z < 3.5$ , from a variability-selected QSO sample. For higher redshifts ( $z > 3.5$ ), our method appears to be more efficient selecting fainter QSOs ( $g > 22.5$ ), due to the intrinsic incompleteness of the SDSS data at such magnitudes (the completeness of SDSS imaging data rapidly decays for  $r > 22$ ; Aihara et al. 2011).

### 5.4. Cross-match with spectroscopic surveys

We cross-match our catalog of selected LAE candidates with the catalogs of SDSS DR18 (Almeida et al. 2023), DESI EDR



**Fig. 17.** Differential number counts per square degree as a function of  $\text{Ly}\alpha$  luminosity (top) and  $M_{\text{UV}}$  (bottom). We show the counts for the full sample (dashed red line), the visual golden sample (solid red line) and the subsample with confirmed spectroscopic redshift in either SDSS, HETDEX or DESI (teal solid histograms). Most of the objects in the spectroscopic subsample have matching photometric redshifts ( $\sim 90\%$ ).

(DESI Collaboration 2024) and HETDEX Public Source Catalog 1 (Mentuch Cooper et al. 2023). The match with these catalogs yield a subsample of 370 objects with spectroscopic counterpart: 229, 132 and 9 in SDSS, DESI and HETDEX, respectively. The comparison between our measured redshifts and the spectroscopic redshifts measured by the respective surveys is shown in Fig. 5, and discussed in Sect. 3.2.6 above.

In Fig. 17 we show the differential number counts per unit area in our selected sample as compared to the subsample having spectroscopic counterparts. While the overlap between the coverage of SDSS DR16 and all three PAUS wide fields is total, there is only partial overlap between those and HETDEX Public Catalog 1 and DESI EDR. Our sample is only slightly more complete than that of SDSS DR16 for  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \gtrsim 44.25$ , and contains a significantly larger number of objects for fainter  $\text{Ly}\alpha$  luminosities ( $\sim 1$  dex larger at  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \approx 43.75$ ). On the other hand, the ratio between the number density of our sample and the QSO sample in the DESI EDR appears to be constant down to  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \approx 43.5$ .

## 6. Summary

In this work, we present a method to select QSOs through  $\text{Ly}\alpha$  emission lines in the Physics of the Accelerating Universe Survey (PAUS) filter system. This method is a refinement of that developed in Torralba-Torregrosa et al. (2023). We built mock catalogs of the target sample, QSOs with strong  $\text{Ly}\alpha$  emission,

and the expected contaminants: that is QSOs at redshifts where Ly $\alpha$  cannot be observed with our NB filters, and low- $z$  galaxies.

We probed a total effective area of 35.68 deg<sup>2</sup> in three PAUS wide fields (PAUS W1, W2, and W3), and obtained a sample of 915 Ly $\alpha$ -selected QSOs in the redshift range  $2.7 < z < 5.3$ , with Ly $\alpha$  luminosities of  $43.5 < \log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) < 45.5$ . Our method leads to the obtention of a sample of candidates selected through prominent flux excess in one or more NBs, which can be compatible with QSO lines at such redshifts. We then used a neural-network classifier trained with mock catalogs to obtain a purer sample. This classifier is able to correctly pick up Ly $\alpha$ -emitting QSOs from our preselected sample with an accuracy of  $\approx 95\%$  while maintaining a low false-positive rate ( $\approx 7\%$ ).

A random forest regressor was trained with the same mock catalogs and was used to obtain precise redshifts. The determination of the redshift using the Ly $\alpha$  detection NB only is slightly inaccurate, because the intrinsic width of the observed wavelength coverage of the NBs corresponds to a redshift uncertainty of  $\Delta z \approx 0.12$ . Moreover, as QSO lines can be fairly wide in velocity space (up to  $\sim 5000 \text{ km s}^{-1}$ ), the pivot wavelength of the selected NB could not exactly correspond to the peak of the Ly $\alpha$  line. In addition, there is an expected bias of the QSO Ly $\alpha$  redshift with respect to the systemic redshift. Our algorithm effectively obtains redshifts that are more precise than those measured spectroscopically.

The purity of our sample reaches  $\approx 90\%$  at  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \approx 44.5$  for the lowest redshift intervals ( $2.75 < z < 3.2$ ) and  $\approx 80\%$  at  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) \approx 44$  for  $3.2 < z < 4.2$ , as estimated using our mocks. The purity of the subsample with spectroscopic counterparts (cross-matching with SDSS, DESI, and HETDEX) is  $\sim 88\%$  for the full selected sample, and  $\approx 91\%$  ( $\approx 100\%$ ) for Ly $\alpha$  luminosity  $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) > 44$  ( $\log_{10}(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) > 44.5$ ). The main source of contamination, as predicted from our mocks and confirmed by the spectroscopic cross-matched subsample, is the lower- $z$  QSOs ( $z < 2.7$ ). In particular, the strong C IV QSO line is the main source of misidentification with Ly $\alpha$ .

We estimated the Ly $\alpha$  and UV ( $M_{\text{UV}}$ , at rest-frame 145 nm) LFs. For this task, we obtained estimations of purity and number-count corrections over a grid of  $r_{\text{synth}}$  (a synthetic broad-band from the combination of NBs, mimicking the classical band  $r$ ) and  $L_{\text{Ly}\alpha}$ , in different intervals of redshift. Both the Ly $\alpha$  and UV LFs show evident evolution, with declining normalization parameter  $\Phi^*$ , fitting a Schechter function and a double power law, respectively. The faint end of the Ly $\alpha$  LF shows a clear correlation with redshift in the probed interval, suggesting that the number density of bright quasars declines faster than that of the faint counterparts, in agreement with the SMBH growth models at  $z \gtrsim 2$ .

We obtained composite QSO spectra from NB data for different redshift intervals. We measured a UV continuum slope of  $\beta_{\text{UV}} = -1.55 \pm 0.02$ . We employ the composite spectra to measure the mean IGM transmission of the Ly $\alpha$  forest, obtaining results compatible with previous spectroscopic determinations. The total Ly $\alpha$  luminosity density obtained from the integral of the Ly $\alpha$  LF shows a rapid decline with increasing redshift, being subdominant by several orders of magnitude already by  $z \approx 4$ , with respect to that of star-formation Ly $\alpha$  emission.

## Data availability

The catalog of selected objects elaborated in this work is available at [CosmoHub](https://cosmohub.pic.es) (Tallada et al. 2020; Carretero et al. 2017), together with the whole PAUS public data release, through this link: <https://cosmohub.pic.es/catalogs/319>

The data are freely accessible after creating an account at <https://cosmohub.pic.es/register>. More information on the data release, as well as access to the PAUS database and raw and reduced images, can be found at <https://pausurvey.org/public-data-release/>. The stacking code `stonp` is publicly available at <https://github.com/PAU-survey/stonp>

*Acknowledgements.* The authors acknowledge the insightful feedback from the scientific referee, which has significantly contributed to enhancing the quality of this paper. This work has been funded by project PID2019-109592GB-I00/AEI/10.13039/501100011033 from the Spanish Ministerio de Ciencia e Innovación (MCIN)–Agencia Estatal de Investigación, by the Project of Excellence Prometeo/2020/085 from the Conselleria d’Innovació Universitat, Ciència i Societat Digital de la Generalitat Valenciana. The authors acknowledge the financial support from the MCIN with funding from the European Union NextGenerationEU and Generalitat Valenciana in the call Programa de Planes Complementarios de I+D+i (PRTR 2022). Project (VAL-JPAS), reference ASFAE/2022/025. PR and DS acknowledge the support by the Tsinghua Shui Mu Scholarship. PR, DS and ZC acknowledge the funding of the National Key R&D Program of China (grant no. 2023YFA1605600), the science research grants from the China Manned Space Project with No. CMS-CSST2021-A05, and the Tsinghua University Initiative Scientific Research Program (No. 20223080023). PR acknowledges additional funding from the National Science Foundation of China (grant no. 12350410365). EG acknowledges grants from Spain Plan Nacional (PGC2018-102021-B-100) and Maria de Maeztu (CEX2020-001058-M). The PAU Survey is partially supported by MINECO under grants CSD2007-00060, AYA2015-71825, ESP2017-89838, PGC2018-094773, PGC2018-102021, PID2019-111317GB, SEV-2016-0588, SEV-2016-0597, MDM-2015-0509 and Juan de la Cierva fellowship and LACEGAL and EWC Marie Skłodowska-Curie grant No. 734374 and no. 776247 with ERDF funds from the EU Horizon 2020 Programme, some of which include ERDF funds from the European Union. IEBC and IFAE are partially funded by the CERCA and Beatriu de Pinos program of the Generalitat de Catalunya. Funding for PAUS has also been provided by Durham University (via the ERC StG DEGAS-259586), ETH Zurich, Leiden University (via ERC StG ADULT-279396 and Netherlands Organisation for Scientific Research (NWO) Vici grant 639.043.512), University College London and from the European Union’s Horizon 2020 research and innovation programme under the grant agreement No. 776247 EWC. The PAU data center is hosted by the Port d’Informació Científica (PIC), maintained through a collaboration of CIEMAT and IFAE, with additional support from Universitat Autònoma de Barcelona and ERDF. We acknowledge the PIC services department team for their support and fruitful discussions. EG acknowledges grants from Spain Plan Nacional (PGC2018-102021-B-100) and Maria de Maeztu (CEX2020-001058-M). JC acknowledges support from the grant PID2021-123012NA-C44 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”. FJC is supported by grants PID2022-141079NB and CEX2020-001058-M funded by MCIN/AEI/10.13039/501100011033. ME acknowledges funding by MCIN with funding from European Union NextGenerationEU (PRTR-C17.11) and by Generalitat de Catalunya. IFT participation is supported by the grant PID2021-123012NB-C43P funded by MCIN/AEI/10.13039/501100011033. H. Hildebrandt is supported by a DFG Heisenberg grant (Hi 1495/5-1), the DFG Collaborative Research Center SFB1491, as well as an ERC Consolidator Grant (No. 770935). IFAE participation is supported by grant PID2021-123012NB funded by MCIN/AEI/10.13039/501100011033. CIEMAT participation is supported by the grant PID2021-123012NB-C42P funded by MCIN/AEI/10.13039/501100011033.

## References

- Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, *ApJS*, **192**, 5  
 Adams, N. J., Bowler, R. A. A., Jarvis, M. J., Varadaraj, R. G., & Häufler, B. 2023, *MNRAS*, **523**, 327  
 Aihara, H., Allende Prieto, C., An, D., et al. 2011, *ApJS*, **193**, 29  
 Akins, H. B., Casey, C. M., Lambrides, E., et al. 2024, *ApJ*, submitted [arXiv:2406.10341]  
 Akiyama, M., He, W., Ikeda, H., et al. 2018, *PASJ*, **70**, S34  
 Alarcon, A., Gaztanaga, E., Eriksen, M., et al. 2021, *MNRAS*, **501**, 6103  
 Almeida, A., Anderson, S. F., Argudo-Fernández, M., et al. 2023, *ApJS*, **267**, 44  
 Arrabal Haro, P., Rodríguez Espinosa, J. M., Muñoz-Tuñón, C., et al. 2020, *MNRAS*, **495**, 1807  
 Askar, A., Davies, M. B., & Church, R. P. 2022, *MNRAS*, **511**, 2631  
 Bacon, R., Brinchmann, J., Richard, J., et al. 2015, *A&A*, **575**, A75  
 Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, *MNRAS*, **430**, 2067

- Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, *ApJ*, 736, 31
- Bonoli, S., Calderone, G., Abramo, R., et al. 2021, *Proc. Int. Astron. Union*, 356, 12
- Boyle, B. J., Shanks, T., Croom, S. M., et al. 2000, *MNRAS*, 317, 1014
- Byrohl, C., & Nelson, D. 2023, *MNRAS*, 523, 5248
- Calhau, J., Sobral, D., Santos, S., et al. 2020, *MNRAS*, 493, 3341
- Carretero, J., Tallada, P., Casals, J., et al. 2017, *Proceedings of the European Physical Society Conference on High Energy Physics*, 5–12 July, 488
- Cenarro, A. J., Moles, M., Cristóbal-Hormillos, D., et al. 2019, *A&A*, 622, A176
- Chabrier, G. 2003, *PASP*, 115, 763
- Compostella, M., Cantalupo, S., & Porciani, C. 2014, *MNRAS*, 445, 4186
- Cowie, L. L., & Hu, E. M. 1998, *AJ*, 115, 1319
- Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, *MNRAS*, 399, 1755
- Cuillandre, J.-C. J., Withington, K., Hudelot, P., et al. 2012, *SPIE Conf. Ser.*, 8448, 84480M
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, *MNRAS*, 486, 2827
- Dawson, S., Rhoads, J. E., Malhotra, S., et al. 2007, *ApJ*, 671, 1227
- Dayal, P., Volonteri, M., Choudhury, T. R., et al. 2020, *MNRAS*, 495, 3065
- de La Vieuville, G., Bina, D., Pello, R., et al. 2019, *A&A*, 628, A3
- DESI Collaboration (Adame, A. G., et al.) 2024, *AJ*, 168, 58
- Ding, J., Madau, P., & Prochaska, J. X. 2024, *MNRAS*, 532, 2082
- Drake, A. B., Garel, T., Wisotzki, L., et al. 2017, *A&A*, 608, A6
- Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, *MNRAS*, 513, 439
- Dubois, Y., Peirani, S., Pichon, C., et al. 2016, *MNRAS*, 463, 3948
- Eriksen, M., Alarcon, A., Cabayol, L., et al. 2020, *MNRAS*, 497, 4565
- Fabian, A. C. 2012, *ARA&A*, 50, 455
- Fan, X. 1999, *AJ*, 117, 2528
- Fanidakis, N., Baugh, C. M., Benson, A. J., et al. 2012, *MNRAS*, 419, 2797
- Faucher-Giguère, C.-A., Prochaska, J. X., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008, *ApJ*, 681, 831
- Gebhardt, K., Mentuch Cooper, E., Ciardullo, R., et al. 2021, *ApJ*, 923, 217
- Glikman, E., Bogosavljević, M., Djorgovski, S. G., et al. 2010, *ApJ*, 710, 1498
- Glikman, E., Djorgovski, S. G., Stern, D., et al. 2011, *ApJ*, 728, L26
- Greene, J. E., Labbe, I., Goulding, A. D., et al. 2024, *ApJ*, 964, 39
- Greig, B., Mesinger, A., McGreer, I. D., Gallerani, S., & Haiman, Z. 2016, *MNRAS*, 466, 1814
- Greig, B., Mesinger, A., McGreer, I. D., Gallerani, S., & Haiman, Z. 2017, *MNRAS*, 466, 1814
- Guaita, L., Gawiser, E., Padilla, N., et al. 2010, *ApJ*, 714, 255
- Hainline, K. N., Shapley, A. E., Greene, J. E., & Steidel, C. C. 2011, *ApJ*, 733, 31
- Harris, D. W., Jensen, T. W., Suzuki, N., et al. 2016, *AJ*, 151, 155
- Hasinger, G., Miyaji, T., & Schmidt, M. 2005, *A&A*, 441, 417
- Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, *MNRAS*, 451, 2663
- Hernán-Caballero, A., Varela, J., López-Sanjuan, C., et al. 2021, *A&A*, 654, A101
- Hernán-Caballero, A., Willmer, C. N. A., Varela, J., et al. 2023, *A&A*, 671, A71
- Hernán-Caballero, A., Akhlaghi, M., López-Sanjuan, C., et al. 2024, *A&A*, 684, A61
- Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, *MNRAS*, 427, 146
- Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, *ApJ*, 654, 731
- Hu, E. M., Cowie, L. L., & McMahon, R. G. 1998, *ApJ*, 502, L99
- Hu, W., Wang, J., Zheng, Z.-Y., et al. 2019, *ApJ*, 886, 90
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, 442, 1805
- Ivashchenko, G., Sergijenko, O., & Torbaniuk, O. 2014, *MNRAS*, 437, 3343
- Izquierdo-Villalba, D., Angulo, R. E., Orsi, A., et al. 2019, *A&A*, 631, A82
- Izquierdo-Villalba, D., Bonoli, S., Dotti, M., et al. 2020, *MNRAS*, 495, 4681
- Izquierdo-Villalba, D., Sesana, A., Bonoli, S., & Colpi, M. 2022, *MNRAS*, 509, 3488
- Jiang, L., McGreer, I. D., Fan, X., et al. 2016, *ApJ*, 833, 222
- Karman, W., Caputi, K. I., Grillo, C., et al. 2015, *A&A*, 574, A11
- Khusanova, Y., Le Fèvre, O., Cassata, P., et al. 2020, *A&A*, 634, A97
- Kocevski, D. D., Finkelstein, S. L., Barro, G., et al. 2024, *ApJ*, submitted [arXiv:2404.03576]
- Kokorev, V., Caputi, K. I., Greene, J. E., et al. 2024, *ApJ*, 968, 38
- Konno, A., Ouchi, M., Shibuya, T., et al. 2018, *PASJ*, 70, S16
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019, *MNRAS*, 488, 1035
- Labbé, I., van Dokkum, P., Nelson, E., et al. 2023, *Nature*, 616, 266
- Laur, J., Tempel, E., Tamm, A., et al. 2022, *A&A*, 668, A8
- Lin, X., Zheng, Z., & Cai, Z. 2022, *ApJS*, 262, 38
- Liu, C., Gebhardt, K., Cooper, E. M., et al. 2022, *ApJ*, 935, 132
- Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, *MNRAS*, 449, 4204
- Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, *ApJS*, 250, 8
- Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, A15
- Malhotra, S., & Rhoads, J. E. 2004, *ApJ*, 617, L5
- Martí, P., Miquel, R., Castander, F. J., et al. 2014, *MNRAS*, 442, 92
- Martin, C. L., & Sawicki, M. 2004, *ApJ*, 603, 414
- Matsuoka, Y., Strauss, M. A., Kashikawa, N., et al. 2018, *ApJ*, 869, 150
- Mathee, J., Sobral, D., Best, P., et al. 2017, *MNRAS*, 471, 629
- Mathee, J., Naidu, R. P., Brammer, G., et al. 2024, *ApJ*, 963, 129
- McGreer, I. D., Jiang, L., Fan, X., et al. 2013, *ApJ*, 768, 105
- Mentuch Cooper, E., Gebhardt, K., Davis, D., et al. 2023, *ApJ*, 943, 177
- Meusinger, H., Hinze, A., & de Hoon, A. 2011, *A&A*, 525, A37
- Morales, A. M., Mason, C. A., Bruton, S., et al. 2021, *ApJ*, 919, 120
- Mountrichas, G. 2023, *A&A*, 672, A98
- Nakajima, K., Fletcher, T., Ellis, R. S., Robertson, B. E., & Iwata, I. 2018, *MNRAS*, 477, 2098
- Navarro-Gironés, D., Gaztañaga, E., Crocce, M., et al. 2023, *MNRAS*, submitted [arXiv:2312.07581]
- Niida, M., Nagao, T., Ikeda, H., et al. 2020, *ApJ*, 904, 89
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
- Ono, Y., Itoh, R., Shibuya, T., et al. 2021, *ApJ*, 911, 78
- Onorato, S., Hennawi, J. F., Schindler, J. T., et al. 2024, *MNRAS*, submitted [arXiv:2406.07612]
- Orsi, Á., Padilla, N., Groves, B., et al. 2014, *MNRAS*, 443, 799
- Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, *ApJS*, 176, 301
- Padilla, C., Castander, F. J., Alarcón, A., et al. 2019, *AJ*, 157, 246
- Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, *A&A*, 530, A122
- Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2016, *A&A*, 587, A41
- Pan, Z., Jiang, L., Fan, X., Wu, J., & Yang, J. 2022, *ApJ*, 928, 172
- Partridge, R. B., & Peebles, P. J. E. 1967, *ApJ*, 147, 868
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825
- Planck Collaboration VI. 2020, *A&A*, 641, A6
- Pritchett, C. J. 1994, *PASP*, 106, 1052
- Queiroz, C., Abramo, L. R., Rodrigues, N. V. N., et al. 2023, *MNRAS*, 520, 3476
- Renard, P., Gaztanaga, E., Croft, R., et al. 2021, *MNRAS*, 501, 3883
- Renard, P., Siudek, M., Eriksen, M. B., et al. 2022, *MNRAS*, 515, 146
- Renard, P., Spinoso, D., Sun, Z., et al. 2024, *ArXiv e-prints* [arXiv:2406.18775]
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, *AJ*, 131, 2766
- Rollinde, E., Theuns, T., Schaye, J., Pâris, I., & Petitjean, P. 2013, *MNRAS*, 428, 540
- Ross, N. P., McGreer, I. D., White, M., et al. 2013, *ApJ*, 773, 14
- Santos, S., Sobral, D., & Mathee, J. 2016, *MNRAS*, 463, 1678
- Santos, S., Sobral, D., Mathee, J., et al. 2020, *MNRAS*, 493, 141
- Santos, S., Sobral, D., Butterworth, J., et al. 2021, *MNRAS*, 505, 1117
- Schechter, P. 1976, *ApJ*, 203, 297
- Schindler, J.-T., Fan, X., McGreer, I. D., et al. 2017, *ApJ*, 851, 13
- Schindler, J.-T., Fan, X., McGreer, I. D., et al. 2019, *ApJ*, 871, 258
- Schmidt, M. 1968, *ApJ*, 151, 393
- Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, *AJ*, 139, 2360
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J. K. 2016, *A&A*, 585, A87
- Serrano, S., Gaztañaga, E., Castander, F. J., et al. 2023, *MNRAS*, 523, 3287
- Shen, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2020, *MNRAS*, 495, 3252
- Shull, J. M., Stevans, M., & Danforth, C. W. 2012, *ApJ*, 752, 162
- Sobral, D., Santos, S., Mathee, J., et al. 2018, *MNRAS*, 476, 4725
- Spinoso, D., Orsi, A., López-Sanjuan, C., et al. 2020, *A&A*, 643, A149
- Spinoso, D., Bonoli, S., Valiante, R., Schneider, R., & Izquierdo-Villalba, D. 2023, *MNRAS*, 518, 4672
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Tallada, P., Carretero, J., Casals, J., et al. 2020, *Astron. Comput.*, 32, 100391
- Thai, T. T., Tuan-Anh, P., Pello, R., et al. 2023, *A&A*, 678, A139
- Torralba-Torregrosa, A., Gurung-López, S., Arnalte-Mur, P., et al. 2023, *A&A*, 680, A14
- Trebtsch, M., Dubois, Y., Volonteri, M., et al. 2021, *A&A*, 653, A154
- Tremmel, M., Ricarte, A., Natarajan, P., et al. 2024, *Open J. Astrophys.*, 7, 26
- Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, *AJ*, 122, 549
- Wang, F., Yang, J., Fan, X., et al. 2021, *ApJ*, 907, L1
- Ward, S. R., Harrison, C. M., Costa, T., & Mainieri, V. 2022, *MNRAS*, 514, 2936
- White, R. L., Becker, R. H., Fan, X., & Strauss, M. A. 2003, *AJ*, 126, 1
- Wisotzki, L. 2000, *A&A*, 353, 861
- Wold, I. G. B., Malhotra, S., Rhoads, J., et al. 2022, *ApJ*, 927, 36
- Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, *ApJ*, 825, 144
- Wyithe, J. S. B., Yan, H., Windhorst, R. A., & Mao, S. 2011, *Nature*, 469, 181
- Xu, C., Smith, A., Borrow, J., et al. 2023, *MNRAS*, 521, 4356
- Yang, J., Wang, F., Wu, X.-B., et al. 2016, *ApJ*, 829, 33
- Yang, G., Brandt, W. N., Alexander, D. M., et al. 2019, *MNRAS*, 485, 3721
- Zhang, Y., Ouchi, M., Gebhardt, K., et al. 2021, *ApJ*, 922, 167

**Table A.1.** NB filters used to select Ly $\alpha$  emitting quasars.

NB ID	Name <sup>a</sup>	5 $\sigma$ magAB <sup>b</sup>	Ly $\alpha$ z ( $\pm 0.06$ ) <sup>c</sup>	# Selected <sup>d</sup>
1	NB455	22.18	2.76	103
2	NB465	22.17	2.84	120
3	NB475	22.27	2.91	88
4	NB485	22.41	2.99	117
5	NB495	22.36	3.08	80
6	NB505	22.30	3.16	66
7	NB515	22.24	3.25	65
8	NB525	22.30	3.32	29
9	NB535	22.27	3.41	32
10	NB545	22.31	3.50	13
11	NB555	22.11	3.58	26
12	NB565	22.26	3.66	23
13	NB575	22.31	3.73	20
14	NB585	22.13	3.82	25
15	NB595	22.06	3.90	17
16	NB605	22.25	3.98	4
17	NB615	22.34	4.06	8
18	NB625	22.21	4.15	10
19	NB635	22.23	4.23	4
20	NB645	22.39	4.31	2
21	NB655	22.38	4.39	1
22	NB665	22.38	4.47	6
23	NB675	22.35	4.56	0
24	NB685	22.14	4.64	1
25	NB695	22.12	4.72	2
26	NB705	22.27	4.81	1
27	NB715	22.37	4.89	2
28	NB725	22.04	4.97	0
29	NB735	21.95	5.05	2
30	NB745	22.07	5.13	0
31	NB755	21.93	5.21	1

**Notes.** <sup>a</sup>The name of the filters is composed of “NB” + the approximate effective wavelength in nm. <sup>b</sup>Average limit magnitude at the 5 $\sigma$  level. <sup>c</sup>Interval of Ly $\alpha$  redshifts probed. The amplitude of this interval is defined by the FWHM of the filters. <sup>d</sup>Number of selected Ly $\alpha$  emitting quasars in the final sample by each NB.

## Appendix A: Ly $\alpha$ redshift of the PAUS narrow bands

In Table A.1 we display the list of NBs used to select Ly $\alpha$  emitting candidates, the associated Ly $\alpha$  redshift of each NB and the number of candidates selected.

## Appendix B: ANN classifier details

For the machine learning models used in this work, we use the python library Sci-kit learn (Pedregosa et al. 2011). The ANN classifier is built using the MLPClassifier method. We performed a grid-search of optimal parameters with GridSearchCV, over an extense grid of parameters. The ANN source classifier optimal parameters are:

```
hidden_layer_sizes = (60, 60, 60),
batch_size = 64,
alpha = 1e-4,
learning_rate = 'adaptive',
max_iter = 1e4,
solver = 'adam'.
```

**Table D.1.** Best-fit Schechter parameters for the Ly $\alpha$  LF.

Redshift	$\log_{10}(\Phi^*/\text{Mpc}^{-3})$	$\log_{10}(L^*/\text{erg s}^{-1})$	$\alpha$
2.76 - 2.99	$-6.48^{+0.24}_{-0.28}$	$44.89^{+0.21}_{-0.19}$	$-1.44^{+0.18}_{-0.16}$
2.91 - 3.16	$-6.63^{+0.20}_{-0.28}$	$44.98^{+0.21}_{-0.17}$	$-1.30^{+0.17}_{-0.17}$
3.08 - 3.32	$-6.81^{+0.23}_{-0.30}$	$45.06^{+0.22}_{-0.20}$	$-1.34^{+0.18}_{-0.17}$
3.25 - 3.50	$-7.00^{+0.38}_{-0.42}$	$45.05^{+0.26}_{-0.28}$	$-1.60^{+0.26}_{-0.30}$
3.41 - 3.66	$-6.80^{+0.45}_{-0.70}$	$44.73^{+0.44}_{-0.31}$	$-1.63^{+0.39}_{-0.34}$
3.58 - 3.82	$-6.80^{+0.43}_{-0.57}$	$44.77^{+0.39}_{-0.31}$	$-1.62^{+0.32}_{-0.26}$
2.76 - 2.99	$-6.62^{+0.08}_{-0.09}$	45 (fixed)	$-1.51^{+0.08}_{-0.09}$
2.91 - 3.16	$-6.65^{+0.09}_{-0.10}$	45 (fixed)	$-1.30^{+0.10}_{-0.11}$
3.08 - 3.32	$-6.72^{+0.10}_{-0.11}$	45 (fixed)	$-1.28^{+0.11}_{-0.12}$
3.25 - 3.50	$-6.90^{+0.13}_{-0.15}$	45 (fixed)	$-1.54^{+0.14}_{-0.16}$
3.41 - 3.66	$-7.19^{+0.18}_{-0.23}$	45 (fixed)	$-1.81^{+0.18}_{-0.22}$
3.58 - 3.82	$-7.13^{+0.13}_{-0.15}$	45 (fixed)	$-1.76^{+0.13}_{-0.14}$

**Table D.2.** Best-fit double power-law parameters for the UV LF.

Redshift	$\log_{10}(\Phi^*/\text{Mpc}^{-3})$	$M^*$	$\beta$	$\gamma$
2.76 - 2.99	$-6.61^{+0.74}_{-0.45}$	$-26.63^{+1.61}_{-0.75}$	$-1.78^{+0.20}_{-0.33}$	$-3.32^{+1.42}_{-0.91}$
2.91 - 3.16	$-6.75^{+0.89}_{-0.47}$	$-26.74^{+2.09}_{-0.81}$	$-1.80^{+0.20}_{-0.29}$	$-3.04^{+1.72}_{-0.78}$
3.08 - 3.32	$-6.66^{+0.69}_{-0.49}$	$-26.66^{+1.74}_{-0.94}$	$-1.70^{+0.19}_{-0.28}$	$-2.81^{+1.64}_{-0.60}$
3.25 - 3.50	$-6.83^{+0.68}_{-0.47}$	$-26.64^{+1.76}_{-0.92}$	$-1.65^{+0.22}_{-0.30}$	$-2.91^{+1.69}_{-0.72}$
3.41 - 3.66	$-6.83^{+0.44}_{-0.44}$	$-26.49^{+1.09}_{-0.87}$	$-1.54^{+0.24}_{-0.30}$	$-3.29^{+1.47}_{-0.96}$
3.58 - 3.82	$-6.92^{+0.58}_{-0.41}$	$-26.69^{+1.52}_{-0.81}$	$-1.60^{+0.23}_{-0.30}$	$-3.15^{+1.62}_{-0.89}$
2.76 - 2.99	$-5.92^{+0.21}_{-0.13}$	$-25.23^{+0.65}_{-0.39}$	-1.25 (fixed)	$-2.68^{+0.28}_{-0.27}$
2.91 - 3.16	$-5.92^{+0.43}_{-0.23}$	$-24.89^{+1.30}_{-0.71}$	-1.25 (fixed)	$-2.44^{+0.35}_{-0.25}$
3.08 - 3.32	$-6.00^{+0.38}_{-0.21}$	$-25.06^{+1.24}_{-0.68}$	-1.25 (fixed)	$-2.37^{+0.35}_{-0.24}$
3.25 - 3.50	$-6.31^{+0.32}_{-0.17}$	$-25.47^{+1.07}_{-0.56}$	-1.25 (fixed)	$-2.52^{+0.45}_{-0.34}$
3.41 - 3.66	$-6.58^{+0.17}_{-0.12}$	$-26.00^{+0.36}_{-0.38}$	-1.25 (fixed)	$-3.03^{+1.09}_{-0.56}$
3.58 - 3.82	$-6.54^{+0.25}_{-0.16}$	$-25.89^{+0.85}_{-0.55}$	-1.25 (fixed)	$-2.76^{+0.76}_{-0.45}$

In Fig. B.1 we show the confusion matrices of the three classes, in different bins of measured Ly $\alpha$  luminosity.

For the RF redshift regressor we use the method RandomForestRegressor, and perform a grid-search using the same method employed for the NN. We obtain the following optimal parameters:

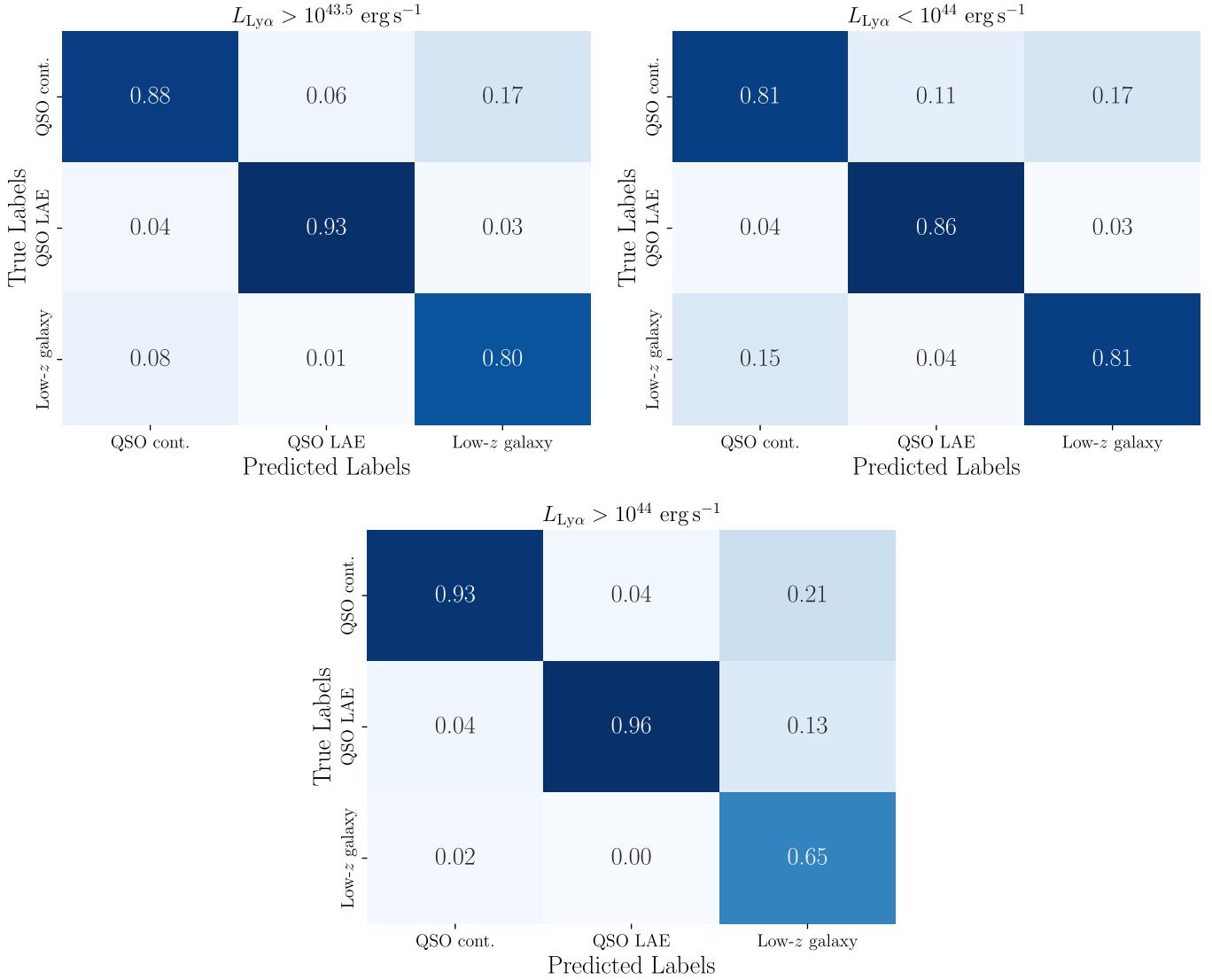
```
min_samples_split = 3,
min_samples_leaf = 2,
max_depth = 20,
bootstrap = True
```

## Appendix C: 2D purity and completeness

In Fig. C.1 we display the purity and completeness of the selection method in bins of  $r$  and measured  $L_{\text{Ly}\alpha}$ . These 2D maps are obtained by applying the selection method on the mocks (Sect. 3.1) and comparing the selected sample with the parent sample using Eqs. 4 and 5.

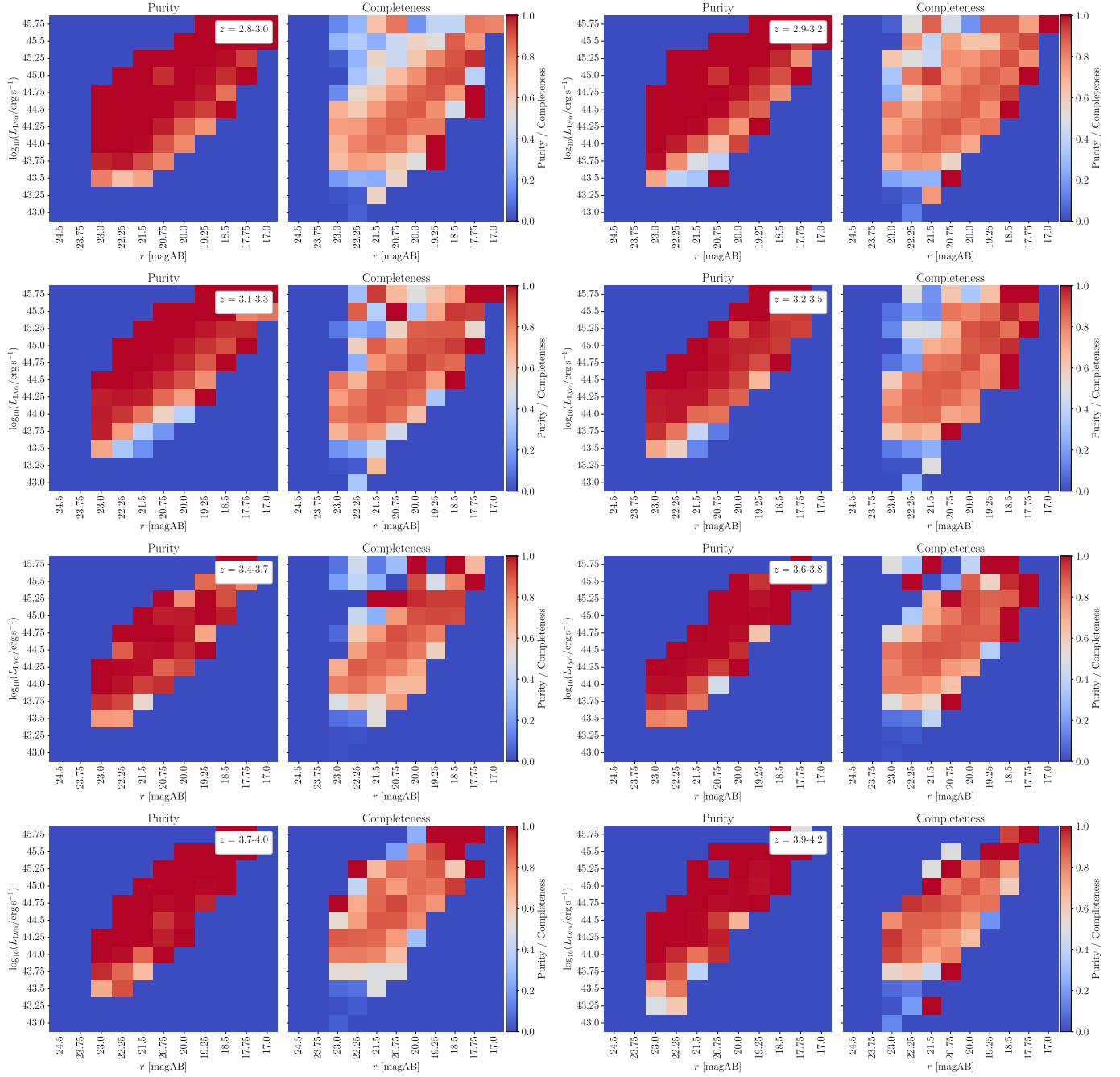
## Appendix D: Schechter and double power-law fit parameters

In Tables D.1 and D.2 we list the best-fit parameters to a Schechter function for the Ly $\alpha$  LF, and double powerlaw for the UV LF, respectively. For the Ly $\alpha$  LF Schechter fit we list the results of fitting with all three free Schechter parameters, and



**Fig. B.1.** Confusion matrix for the ANN classifier used to select LAEs. The three classes considered are: QSOs selected by a feature other than  $\text{Ly}\alpha$  (“QSO cont.”), QSOs correctly selected by their  $\text{Ly}\alpha$  line, in the correct redshift (“QSO LAE”), and low- $z$  ( $z \approx 0\text{--}2$ ) galaxy contaminants (“Low- $z$  galaxy”). The annotations in the cells are the normalized counts so the vertical columns add up to 1. We show the confusion matrices for the full test set (top,  $L_{\text{Ly}\alpha} > 10^{43.5} \text{ erg s}^{-1}$ ) and splitting in low (middle,  $L_{\text{Ly}\alpha} < 10^{44} \text{ erg s}^{-1}$ ) and high luminosity bins (bottom,  $L_{\text{Ly}\alpha} > 10^{44} \text{ erg s}^{-1}$ ). This figure shows that the ANN effectively selects QSO LAEs with an accuracy of 93% for the general sample, while maintaining a low false negative rate (6% and 1% in the case of objects labeled as QSO cont. and low- $z$  galaxies, respectively).

fixing  $\log_{10}(L^*/\text{erg s}^{-1}) = 45$ , as discussed in Sect. 4.2.1. Similarly, for the UV LF fit, we list the results of leaving all four DPL parameters free, and fixing the faint-end slope  $\beta = -1.25$ , as discussed in Sect. 4.3.1



**Fig. C.1.** 2D maps of purity and completeness in bins of  $L_{Ly\alpha}$  and magnitude  $r$ . The values displayed in Fig. 6 are obtained by collapsing these 2D maps over the horizontal axis.