Revealing new high-redshift quasar populations through Gaussian mixture model selection

J. D. Wagenveld[1,2], A. Saxena[3], K. J. Duncan[4,2], H. J. A. Röttgering[2], and M. Zhang[2]

[1] Max-Planck Institut fur Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[2] Leiden Observatory, Leiden University, PO Box 9513, NL 2300 RA Leiden, The Netherlands
[3] Department of Physics and Astronomy, University College London , Gower Street, London, WC1E 6BT, UK
[4] SUPA, Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK

ABSTRACT

We present a novel method for identifying candidate high-redshift quasars (HzQs; ($z \gtrsim 5.5$)) —which are unique probes of supermassive black hole growth in the early Universe— from large-area optical and infrared photometric surveys. Using Gaussian mixture models (GMMs) to construct likelihoods and incorporate informed priors based on population statistics, our method uses a Bayesian framework to assign posterior probabilities that differentiate between HzQs and contaminating sources. We additionally include deep radio data to obtain informed priors. Using existing HzQ data in the literature, we set a posterior threshold that accepts ~90% of known HzQs while rejecting > 99% of contaminants such as dwarf stars or lower redshift galaxies. Running the probability selection on test samples of simulated HzQs and contaminants, we find that the efficacy of the probability method is higher than traditional colour cuts, decreasing the fraction of accepted contaminants by 86% while retaining a similar fraction of HzQs. As a test, we apply our method to the Pan-STARRS Data Release 1 (PS1) source catalogue within the HETDEX Spring field area on the sky, covering 400 sq. deg. and coinciding with deep radio data from the LOFAR Two-metre Sky Survey Data Release 1 (LoTSS DR1). From an initial sample of ~5×10^5 sources in PS1, our selection shortlists 251 candidate HzQs, which are further reduced to 63 after visual inspection. Shallow spectroscopic follow-up of 13 high-probability HzQs resulted in the confirmation of a previously undiscovered quasar at $z = 5.66$ with photometric colours $i - z = 1.4$, lying outside the typically probed regions when selecting HzQs based on colours. This discovery demonstrates the efficacy of our probabilistic HzQ selection method in selecting more complete HzQ samples, which holds promise when employed on large existing and upcoming photometric data sets.

Key words. (Galaxies:) quasars: supermassive black holes – Galaxies: high-redshift – Methods: statistical

1. Introduction

Studying large statistical samples of high-redshift quasars (HzQs) is essential for understanding the formation and evolution of super-massive black holes (SMBHs) in the early Universe. The presence of Gunn-Peterson (GP) troughs (Gunn & Peterson 1965) in the spectra of HzQs at $z \sim 6$, which are caused by near-complete absorption of Lyα photons by the increasingly neutral intergalactic medium (IGM) along the line of sight, make them crucial probes of cosmic reionisation (EoR; Fan 2006; Becker et al. 2015). These GP troughs can in turn be used to photometrically identify large samples of HzQs, and the proliferation of wide-area multi-band photometric surveys at optical wavelengths, such as the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2003) and the Panoramic Survey Telescope and Rapid Response System surveys (Pan-STARRS; Chambers et al. 2016), has enabled the discovery of statistically significant samples of bright quasars at high redshifts, with now over 500 confirmed HzQs at $z > 5$ (see Ross & Cross 2020, for a compilation).

For HzQs at $z \sim 6$, towards the end of the EoR, the GP trough falls between the i- and z-band filters. Therefore, in the context of the SDSS and Pan-STARRS surveys (carried out using the u, g, r, i, z, and y filters), quasars at $z \sim 6$ may be identified through pin-pointing 'i-dropout' sources, that show extreme $i - z$ colours. Searches for HzQs using photometric dropout techniques over large areas of the sky often employ linear cuts in magnitude and colours (e.g. Bañados et al. 2015, 2016). For example, a colour cut of $i - z > 1.5$ to 2.5 is typically implemented in addition to magnitude cuts to ensure a balance between the selection of robustly detected HzQs and the exclusion of as many contaminating foreground sources as possible, which are often M-, L-, and T-type brown dwarf stars in the Milky Way (Fan et al. 2001; Willott et al. 2005; Bañados et al. 2016; Jiang et al. 2016).

A radio detection can considerably aid in removing foreground contaminants such as dwarf stars that do not emit persistent radio continuum emission at the sensitivity of current observations (e.g. Burningham et al. 2016), as around 10% of HzQs are seen to be 'radio-loud' (radio loudness being defined as the ratio between rest frame radio and optical flux density) even out to high redshifts (e.g. Bañados et al. 2015). However, overlapping deep radio data are often not available for the large sky areas covered by optical surveys from which candidate HzQs are selected. Deep radio continuum surveys of large sky areas, such as the Low-Frequency Array (LOFAR) Two Metre Sky Survey (LoTSS; Shimwell et al. 2019), can therefore potentially provide valuable additional information that could help improve HzQ selection and minimise the probability of contaminants.

While selection based on optical and infrared colours from large-area surveys has been highly successful in identifying some of the most distant HzQs currently known (e.g. Fan et al.

2001; Willott et al. 2010; Bañados et al. 2016; Matsuoka et al. 2016, 2018; Pipien et al. 2018; Reed et al. 2019), the use of linear cuts may lead to potential biases in the samples of HzQ candidates. A binary cut in colour and magnitude may inevitably lead to a loss of promising HzQ candidates. Additionally, Mortlock et al. (2012) argued that linear cuts result in uniform grouping of high-S/N candidates with more marginal ones that lie near the edges of the selection region, possibly making spectroscopic follow-up harder to prioritise. Finally, HzQs lying close to the limits of the redshift ranges probed by colour selections may be missed due to the GP trough not being fully sampled by the relevant broadband filters used for dropout selection. For example, at redshifts of $z \sim 5.5$, the i-dropout selection may result in certain sources being missed, possibly presenting a gap in our understanding of SMBH evolution and/or the later stages of cosmic reionisation (Yang et al. 2017).

Additionally, binary selection criteria are often unable to properly account for the observational uncertainties in the observed properties for either individual sources or the population of sources being targeted. To overcome these limitations specifically in the case of identifying HzQs, a probabilistic selection as opposed to a binary selection may represent a better way to both obtain more complete samples of HzQs and assign higher probabilities for spectroscopic follow-up to more promising candidates. Bovy et al. (2011) introduced an implementation of Gaussian mixture modelling (GMM) that assumes and then deconvolves a model of the underlying population of sources from data, leading to a robust estimate of probability distribution of sources such as HzQs even from noisy measurements. Such an approach has been successfully employed to assign probabilities and better select low- and intermediate-redshift quasar candidates from SDSS data (e.g. Bovy et al. 2011, but see also Bailer-Jones et al. 2008; Richards et al. 2009).

Further complexities can be introduced in these models to improve the probability assignment, for example by also taking into account the respective prior probabilities of the different contaminants – particularly dwarf stars in the Galaxy – based on their spatial distribution and number densities on the sky. Such an approach was implemented by Mortlock et al. (2012) for HzQs where prior information about populations that exhibit HzQ-like colours was used to assign a contamination (and as a result HzQ) probability and reduce the number of contaminants, leading to the discovery of a quasar at $z = 7.1$ (Mortlock et al. 2011). However, the initial selection of HzQ candidates in the probabilistic approach of Mortlock et al. (2012) still relied on linear colour cuts, and could potentially suffer from the same incompleteness issues as faced by other colour-based HzQ selections.

Therefore, there remains room for improvement in probabilistic HzQ selection methods, namely by more accurately constraining the luminosity and sky distributions of possible contaminants to obtain more complete samples of HzQs. In this work, we build upon the probabilistic approach of selecting HzQs based on posterior probability estimation using informed priors and likelihood estimation utilising GMMs. We also we make use of deep radio observations of the HETDEX spring field taken as part of the LoTSS first data release (LoTSS DR1; Shimwell 2017), using radio detection as an additional prior to minimise foreground contamination. With the combination of multi-wavelength data and a probabilistic approach, we aim to develop a selection technique capable of uncovering more complete samples of HzQs from large-area surveys while minimising the number of contaminating sources present in them.

This paper is organised as follows. In Section 2 we describe the data sets that are used, and our HzQ selection method is described Section 3. In Section 4 we apply our selection method to the data sets, obtaining probabilistically selected HzQ candidates. In Section 5 we present spectroscopic follow-up for a handful of identified high-priority HzQ candidates, and report the discovery of a previously undiscovered quasar at $z \sim 5.7$. In Section 6, we discuss the performance of our selection method, application to incoming large sky survey data sets, and the possible implications of the discovery of P144+50. Finally, in Section 7 we summarise the findings of this study.

Throughout the paper, we assume a Planck 2015 cosmology (Planck Collaboration 2015), with $H_0 = 67.8$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.308$, and $\Omega_\Lambda = 0.692$. All magnitudes are given in the AB system (Oke & Gunn 1983) unless otherwise stated.

2. Data

2.1. Pan-STARRS

The primary data set used to identify HzQ candidates in this study is Pan-STARRS Data Release 1 (PS1). The PS1 survey covers 3π steradian of the sky, including the entire northern hemisphere (Chambers et al. 2016), reaching 5σ depths of 23.3, 23.2, 23.1, 22.3, and 21.3 AB in the *g, r, i, z,* and y optical filters, respectively.

We first retrieve a sample of sources from the PS1 data archive[1], and although no colour cuts are made for the initial selection, a number of other criteria are applied to reduce the full PS1 sample down to the appropriate parameter space and more manageable numbers. As we are primarily interested in HzQs, we require a nondetection in the g and r filters while requiring a robust detection in the i, z, and y filters. The nondetections are attributed to magnitudes fainter than the 5σ limiting magnitudes in the photometric filters published by the PS1 team, or values of -999 as this value is the magnitude assigned in case of a nondetection in a particular band.

As a proof of concept, we also restrict our analysis to the sky area corresponding to the HETDEX Spring field, which is advantageously covered by deep radio data at 150 MHz from LoTSS DR1 (see Section 2.2). The selection criteria for obtaining an initial sample from the PS1 catalogue can be summarised as follows:

$160° <$ R.A. $< 232°$

$42° <$ Dec. $< 62°$

$r_{P1} > 23.2$ OR $r_{P1} = -999$

$g_{P1} > 23.2$ OR $g_{P1} = -999$

$i_{P1} > 0, \; z_{P1} > 0, \; y_{P1} > 0.$

Furthermore, objects with flags from the PS1 processing pipeline indicating poor or low-quality detections are excluded to remove sources that have poor photometric data, following Table 6 of Bañados et al. (2014). From these criteria, a sample of $\sim 5 \times 10^5$ sources with complete photometric data is retrieved.

2.2. Radio data

The radio data used in this study are taken from LoTSS DR1 (Shimwell et al. 2017), which is a low-frequency radio continuum survey covering over 424 sq.deg of the northern hemisphere that reaches a median sensitivity of 71μJy beam^{-1}. Consequently, radio sources that are considered are based on the 5σ

---

[1] https://panstarrs.stsci.edu

LoTSS DR1 catalogues and have a flux density of at least 350 μJy. Full details about the data reduction, processing, final images, and source catalogue creation are presented in Shimwell et al. (2019), with robust optical cross identifications presented in Williams et al. (2019) and accompanying photometric redshifts in Duncan et al. (2019). We additionally make use of early LOFAR 'deep-fields' data in the Boötes field (Williams et al. 2018), which coincides with the HETDEX Spring field sky area.

Visual inspection of the radio and optical images of initial HzQ candidate samples drawn from PS1 demonstrated that dusty red galaxies at intermediate redshifts ($0 \lesssim z \lesssim 3$) represented an additional potential contaminant population that could also emit significant radio continuum emission. The Boötes deep field data from Williams et al. (2018) therefore acts as a primary reference for those sources where the high-quality optical data enable robust photometric redshifts.

2.3. Ancillary data

Several other data sets are utilised for training, testing, or validation of the GMM algorithms implemented in this work, or in accurate construction of priors. As a validation sample for HzQs, we use the sample of confirmed HzQs compiled by Bañados et al. (2016) containing all $z > 5.6$ quasars known up to March 2016.

As a reference data set for dwarf stars in the Milky Way, we use a catalogue of brown dwarfs observed in PanSTARRS by Best et al. (2017). From the same work we use data on the mean absolute magnitudes of different dwarf types in PS1, which are also used for constructing their sky densities.

3. High-redshift quasar selection

Having outlined the data sets used to implement our new HzQ selection, in this section we describe the ingredients that go into the construction and implementation of our GMM-based HzQ selection method.

Our HzQ selection method builds upon probabilistic selection of HzQs using a Bayesian framework presented by Mortlock et al. (2012), which does not rely on binary colour–magnitude cuts and incorporates additional prior knowledge about quasars and other contaminants to predict the likelihood of a source selected from a large area survey being a HzQ. For HzQs at $z \gtrsim 5.6$, therefore, a flexible algorithm can be constructed that can compute the probability for any given source based on its i_{P1}, z_{P1}, and y_{P1} magnitudes.

We begin by first defining the posterior distributions for the classes of objects that are likely to occupy the photometric parameter spaces typically occupied by HzQs. We recall that the posterior probability of a source being part of any particular population can be calculated using Bayes' theorem:

$$P(C_k|X) = \frac{P(C_k)\mathcal{L}(X|C_k)}{\sum_{i=1}^{N} P(C_i)P(X|C_i)},\tag{1}$$

where $P(C_k)$ is the prior probability of an object belonging to class k, and $\mathcal{L}(X|C_k)$ is the likelihood of the given source being part of class k, normalised over all N possible classes and their associated probabilities.

In reality, when considering the measurements of astronomical objects, additional factors related to both the distribution of sources on the sky and the survey limitations must be accounted for when deriving probability estimates. More generally, considering these additional factors and the features $\mathbf{f} = \{f_1, f_2, ..., f_n\}$

of a source that differentiate it from other sources in the data, the prior probability of sources belonging to class k with parameters θ_k is calculated as:

$$P(C_k|\mathbf{f}, \det) = \int \rho(\theta_k)P(\det|\theta_k, C_k)d\theta_k ,\tag{2}$$

where $\rho(\theta_k)$ is the sky density, and $P(\det|\theta_k, C_k)$ is the probability that the source is detected in the survey.

For sources detected in a flux-limited survey, the parameters θ_k most relevant to the probability are the magnitudes of the source classes in different filters, described by $\mathbf{m}_k$. In this case the features, $\mathbf{f}$, would describe the observed magnitudes of a given source in different filters, $\hat{\mathbf{m}}$. Therefore, to calculate the prior we marginalise over the relevant magnitude space. The prior is then combined with the likelihood in the 'weighted evidence' term, describing the evidence that the source in question belongs to a given class:

$$W_k(\hat{\mathbf{m}}, \det) = \int \rho(\mathbf{m}_k)P(\det|\mathbf{m}_k, C_k)\mathcal{L}(\hat{\mathbf{m}}|\mathbf{m}_k, C_k)d\mathbf{m}_k ,\tag{3}$$

where $\mathcal{L}(\hat{\mathbf{m}}|\mathbf{m}_k, C_k)$ is the likelihood of the features of a source belonging to an object of class k. With this, Eq. 1 can be rewritten as

$$P(C_k|\hat{\mathbf{m}}, \det) = \frac{W_k(\hat{\mathbf{m}}, \det)}{\sum_{i=1}^{N} W_i(\hat{\mathbf{m}}, \det)}.\tag{4}$$

Having rewritten the equation to calculate the posterior probability of any given class of objects detected in a survey in terms of its apparent magnitude, below we describe the classes of sources that we consider in our search for complete samples of HzQs.

3.1. Source classes

Successful implementation of our HzQ selection method requires the proper identification of all classes of sources that are relevant and overlap with the HzQ parameter spaces. As a result, not every class of astrophysical source needs to be considered, which may also be considered as setting the prior of non-relevant source classes to zero. The relevant classes consist of the target HzQ population and a set of contaminating populations occupying the same feature space. Therefore, we identify three relevant populations: HzQs, dwarf stars within the Milky Way, and dusty intermediate-redshift galaxies with red observed-frame optical colours. To model these populations we require data with PS1 magnitudes for each.

As mentioned above, we use the Galactic brown dwarf stars catalogue from Best et al. (2017) containing photometry of dwarf stars, and the deep multi-wavelength galaxy catalogues in the Boötes field from Williams et al. (2018) containing photometric measurements for dusty intermediate-redshift galaxies. Both catalogues contain ~10^4 sources, which is sufficient to model the colour space reliably without being biased by scatter in individual sources.

However, the same is not true for the catalogue containing approximately 200 confirmed HzQs. Therefore, to model the distribution of the quasar population in the colour spaces probed, we simulate the rest-frame UV spectral energy distributions (SEDs) for a population of quasars using a distribution of power laws, $\alpha \sim \mathcal{N}(1.30, 0.38)$, following the distribution presented by Cristiani et al. (2016). These power-law SEDs are then combined with emission lines using the SDSS quasar template from Vanden Berk et al. (2001).

To then simulate a population of high-redshift quasars, each simulated quasar spectrum (continuum + emission lines) is redshifted. The redshift is drawn from the the redshift distribution following the co-moving luminosity functions as defined in Mortlock et al. (2012), Eq. (13), in the redshift range of $5.6 < z < 6.5$. A redshift-dependent IGM absorption from Madau (1995) is then applied to simulated spectra, and the spectra are then convolved with the Pan-STARRS photometric filters (using prescriptions built into the SMPY PYTHON package[2]). As we are only interested in obtaining a reasonable distribution in colour space, the results are not dependent on the absolute flux of the quasars. This method of generating quasar spectra results in a reliable distribution of HzQ colours, and to maintain consistency with the number of contaminants available to model, we simulate a total of 10^4 quasars in this manner. This method of simulating quasars rests on the assumption that both the Vanden Berk et al. (2001) template spectrum and power-law distribution from Cristiani et al. (2016) are valid to higher redshifts as well. While beyond the scope of this paper, more reliable samples of quasars could be generated using parametric SED modelling, which can account for intrinsic changes in quasar spectra as a function of luminosity and redshift (e.g. Temple et al. 2021)

3.2. Likelihoods and Gaussian mixture modeling

In order to estimate the likelihood of a source belonging to a certain population that is considered in this study, we model each relevant population in the $i_{P1} - z_{P1}, z_{P1} - y_{P1}$ colour space. To do this, we use GMMs, which assume that the probability density of a population can be described by a finite number of weighted Gaussian functions (Reynolds 2009). Therefore, to obtain a probability density, N Gaussian functions each with mean μ_i and variance Σ_i are given a weight w_i, with the condition that the N weights sum up to unity as follows:

$$p(x) = \sum_{i=1}^{N} w_i \mathcal{N}(\mu_i, \Sigma_i) \,, \tag{5}$$

$$0 \le w_i \le 1, \quad \sum_{i=1}^{N} w_i = 1 \,. \tag{6}$$

The GMM is implemented in a machine learning algorithm, which optimises the parameters using expectation maximisation (Dempster et al. 1977). To estimate the number of components needed to model each population adequately, the Bayesian information criterion (BIC) is used (Wit et al. 2012). This use of machine learning techniques to model various populations in colour space is a deviation from the method presented in Mortlock et al. (2012), and this is where the novelty of our method compared to traditional techniques relying on binary cuts in the colour space is best highlighted. The resulting likelihood of any given source belonging to a population follows directly from the GMM:

$$\mathcal{L}(\hat{\mathbf{m}}|\mathbf{m}_k, C_k) = \sum_{i=1}^{N} w_i \mathcal{N}(\mu_i, \Sigma_i) \,. \tag{7}$$

Furthermore, we use an extension of the classical GMM algorithm which implements extreme deconvolution, XDGMM (Holoien et al. 2017). This implementation is particularly suited for noisy data, as it deconvolves the noisy distribution of the population in order to capture the underlying distribution more accurately. This method thus makes use of the uncertainties in the

data, both for deconvolving the models and to assign likelihood to input data. As the error bars are folded into the covariance matrix of the GMMs, sources with larger uncertainties are assigned lower likelihood. The GMM algorithm is used to model the previously defined populations (quasars, dwarf stars, galaxies) in $i_{P1} - z_{P1}, z_{P1} - y_{P1}$ colour space. The log likelihoods (assuming a constant error in magnitude) of the resulting GMMs are shown in Figure 1, along with the sources used to create the models for each population, in the left plot simulated quasars, and in the middle and right plots the sources from the reference catalogues (Best et al. 2017; Williams et al. 2018, respectively). The Gaussian components for each mixture model are also shown, with four components for HzQs, six for dwarf stars, and one for galaxies. The number of components that minimises the BIC is chosen for each population separately.

3.3. Detection prior

Many sources in our sample have faint magnitudes, extending all the way down to the PS1 detection limit. This makes obtaining accurate detection priors necessary not only to differentiate between real and fake sources, but also to robustly characterise the various populations of sources considered, especially at the faintest magnitudes.

As the fraction of real sources detected as a function of source magnitude in PS1 (Metcalfe et al. 2013) is relatively well described by a sigmoid function, the detection priors we use for PS1 detected sources in this study are calculated as:

$$P(\det|\mathbf{m}_k, C_k) = \left[1 + \exp\left(4.84 \cdot (m_{filt} - m_{filt,1/2} - 0.4s)\right)\right]^{-1} \,, \tag{8}$$

where m_{filt} is the magnitude of a source in one of the Pan-STARRS filters, $m_{filt,1/2}$ is the 50% magnitude depth of this filter, and s is a binary value that depends on the source type:

$$s = \begin{cases} 0 & \text{if star/point source} \\ 1 & \text{if galaxy/extended source,} \end{cases} \tag{9}$$

which is a relevant statistic for differentiating between point sources and extended sources.

3.4. Radio detection prior

Deep radio continuum data from LoTSS DR1 are used to complement the available optical data from Pan-STARRS, providing radio detections for a subset of the selected sources. To properly account for radio-detected sources, we modify the source classification based on the likelihood of radio detection, which we implement through the inclusion of an additional radio detection parameter, $f_{R,k}$, into the detection prior. Through this radio detection prior, if a radio counterpart in the LoTSS DR1 images of the input source is present, the radio detection is taken into consideration when computing the HzQ posterior probability.

For HzQs, roughly 10% of the quasar population (e.g. Hooper et al. 1995) is 'radio-loud'. This relation seems to hold at higher redshifts, as Bañados et al. (2015) reported a radio-loud fraction of $\sim 10\%$ for $z > 5$ quasars at 1.4 GHz. Recent results from deep LOFAR survey data at lower redshifts suggest that there is no dichotomy between radio-loud and radio-quiet quasars, and that 30% of quasars can be detected by LOFAR surveys (Gurkan et al. 2019). Similar fractions are found in LoTSS DR2 (36% at $> 2\sigma$ significance; Gloudemans et al.

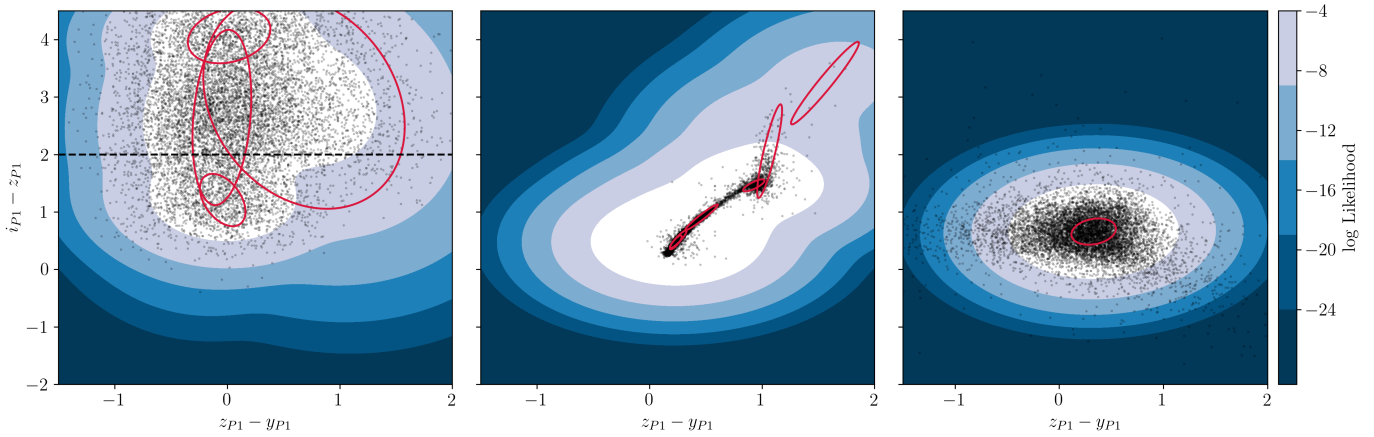[2] https://github.com/dunkenj/smpy

Fig. 1. Log likelihood of quasars (left), dwarf stars (middle), and low-redshift galaxies (right) modelled by a Gaussian mixture model in colour space. The red ellipses indicate the 3-sigma extents of the individual deconvolved Gaussians (four components for quasars, six for dwarf stars, and one for galaxies). The quasar likelihoods are assigned using photometry from simulated quasar spectra as highlighted in Section 3.1, whereas the photometry for dwarf stars and low-redshift galaxies are taken from their respective reference catalogues, Best et al. (2017) and Williams et al. (2018) (black dots). The dashed line in the left panel marks the commonly defined colour cut at $i - z = 2.0$.

2021) at $z \gtrsim 5.0$. A reasonable assumption therefore would be to set $f_{R,k} = 0.3$ for HzQs as the radio detection prior.

For stars, including brown dwarf stars, the radio-loud fraction is very low, and Kimball et al. (2009) find that about one in a million stars can be detected at radio wavelengths. However, low-frequency radio data combined with unparalleled sensitivity from LoTSS represents a new parameter space for the detection of radio signals from stars, as demonstrated by the recent discovery of polarised radio emission from a cold brown dwarf (Vedantham et al. 2020). Nevertheless, bright, non-variable radio continuum emission sufficient to be detected in LoTSS imaging will be significantly rarer for brown dwarf contaminants than for luminous quasars or galaxies. Therefore, the probability of a radio-detected source in our sample being a brown dwarf is virtually zero, with $f_{R,k} = 10^{-6}$ for dwarf stars.

For red, dusty galaxies at intermediate redshift, we find from the deep multi-wavelength catalogues based on deep LOFAR data in the Boötes field (Williams et al. 2018) that only a small fraction (~1%) of these galaxies has a radio detection. Therefore, we set $f_{R,k} = 10^{-2}$ for the galaxy population. We note that these radio detection priors currently represent order of magnitude accuracy, and with deeper data collected over larger areas of the sky by current and future radio surveys, the radio detection priors can be improved upon to further enhance the probability assignment method for HzQs.

3.5. Sky densities

The sky densities of the source classes represent a significant prior, especially given the very rare nature of HzQs that makes any given source on the sky more likely to be a star or foreground galaxy. This prior can also differ depending on the apparent magnitude of the source and in this section we describe the calculation of the sky density priors for source populations considered in this study.

3.5.1. M-,L-, and T-type dwarfs

Since the dwarf star contaminants are all within the Milky Way, the number density of dwarf stars at distance d from the Earth

Table 1. Values of the various parameters used in Eq. (10) with errors as determined by Chen et al. (2001). The Galactic latitude (l) and longitude (b) used here signify the centre of the HETDEX field.

Parameter	Value
$R_{\odot}$	8600 ± 200 pc
$Z_{\odot}$	27 ± 4 pc
h_R	2250 ± 1000 pc
h_Z	330 ± 3 pc
l	120 deg.
b	65 deg.

and Galactic latitude (l) and longitude (b) can be estimated assuming a Galactic model (e.g. Chen et al. 2001) as

$$n_i(d, l, b) = n_{0,i} \exp\left(-\frac{d \cos b \cos l}{h_R}\right) \exp\left(-\frac{|Z_{\odot}| + d \sin b}{h_Z}\right), \quad (10)$$

where $Z_{\odot}$ is the height of the Sun or Earth above the Galactic plane, and h_Z and h_R are the characteristic height and distance scales for stars in the Milky Way, respectively (see also Caballero et al. 2008). The fiducial values of the various parameters used to calculate the sky densities of dwarf stars are given in Table 1.

Given the magnitude range specified, every dwarf type will have a slightly different heliocentric distance at which it will appear in the sample. To calculate this, the absolute PS1 magnitudes of each dwarf type are used from the Best et al. (2017) catalogue. We calculate the sky density for each magnitude bin by integrating the spatial density over the cone covering the sky area. For a single stellar type, this results in

$$\rho_i(\mathbf{m}_i) = \int_{d_1}^{d_2} n_i(D) D^2 dD, \quad (11)$$

where D is the distance in parsecs. The total sky density of all contaminating dwarfs is calculated as the sum of the densities ρ_i of all stellar dwarf types.
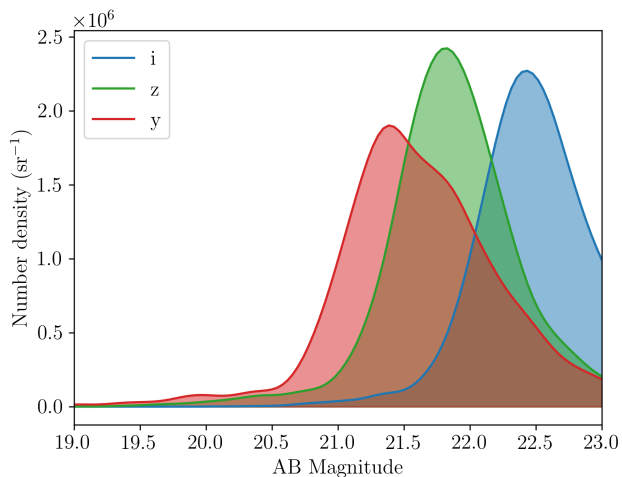
Fig. 2. Expected number density of the galaxy populations as a function of apparent magnitude in PS1 *i*, *z*, and *y* bands extrapolated from the deep multi-wavelength galaxy catalogue in the Boötes field (Williams et al. 2018).

3.5.2. Galaxies

A significant fraction of sources in the PS1 data described in Section 2 are identified as faint red galaxies. As mentioned previously, the information for this population is primarily taken from the Boötes multi-wavelength photometric catalogue from Williams et al. (2018), which also contains photometric redshifts, allowing us to select such galaxies in the redshift range $0 \lesssim z \lesssim 3$. As this is a less well-defined astrophysical population compared to quasars and dwarf stars, we have no luminosity function to model their observed sky density. Instead, we use the apparent magnitudes of these galaxies in the PS1 *i*, *z*, and *y* filters to model their distribution as a function of apparent magnitude in a given filter. We model the distribution with the Kernel density estimation (KDE) technique (Silverman 2017), where we use a bandwidth $h = 0.1$ to get a smooth and continuous representation of the data.

This population of galaxies is made up of the population identified in the Boötes field, selected with the same criteria as the main PS1 sample (see Section 2.1). The Boötes field covers an area of $S \simeq 11.6 \deg^2$ on the sky, and we use this to convert the modelled distribution of galaxies to sky densities. Assuming that the galaxies are isotropically distributed, these sky densities are independent of the direction in which we observe, making the model valid for data in the HETDEX field as well. The resulting sky densities of galaxies detected in the PS1 *i*, *z*, and *y* band data as a function of AB magnitude are shown in Figure 2.

3.5.3. High-redshift quasars

Density functions of HzQs can be expressed in terms of the luminosity functions at high redshifts, which are poorly constrained compared to lower redshifts because of a lack of statistical samples (e.g. Manti et al. 2017). Using observations of quasars across redshifts, Mortlock et al. (2012) derived a redshift- and absolute-magnitude-dependent co-moving luminosity function for HzQs.

In order to calculate absolute magnitudes from the range of observed magnitudes in all relevant PS1 filters, *K*-corrections to the quasar spectra are calculated. We use the method with which we simulated quasar magnitude in Section 3.1, applying

redshift-dependent Lyman absorption from the intervening IGM based on redshift to the quasar SED templates from Vanden Berk et al. (2001). We note that we do not account for the presence of ionised proximity zones around the HzQs. The Lyman absorbed and redshifted spectra are divided by the unaltered SED templates, and convolved with the relevant PS1 filters to obtain the *K*-corrections (following Hogg et al. 2002).

Finally, integrating the redshift- and magnitude-dependent HzQ luminosity density, $\Phi_q(M, z)$, from Mortlock et al. (2012) over the observed redshift cone yields the sky density of HzQs

$$\rho_q(\mathbf{m}_q) = \int_{D_{\mathrm{co},1}}^{D_{\mathrm{co},2}} \Phi_q(M, z) \, D_{\mathrm{co}}^2 \, dD_{\mathrm{co}}, \tag{12}$$

where D_{co} is the co-moving distance in megaparsecs integrated over the distances $(D_{\mathrm{co},1}, D_{\mathrm{co},2})$ corresponding to the redshift range probed.

3.6. Full posterior

For the full photometric sample outlined in Section 2 we calculate the evidences for each class using the priors and likelihoods outlined above. The final quasar posterior probability is then constructed as

$$P_q(\hat{\mathbf{m}}, \det) = \frac{W_q(\hat{\mathbf{m}}, \det)}{\sum_{i=1}^{N} W_i(\hat{\mathbf{m}}, \det)}, \tag{13}$$

where the weighted evidence is calculated based on the priors obtained using the *i*, *z*, and *y* magnitudes and likelihoods in the $i - z$ and $z - y$ colour spaces for each source

$$W_k(\hat{\mathbf{m}}, \det) =$$
$$\int \rho_k(i, z, y) P(\det|i, z, y, C_k) \mathcal{L}(\hat{\mathbf{m}}|i - z, z - y, C_k) \mathrm{d}i \mathrm{d}y \mathrm{d}z . \tag{14}$$

As a result, every source with a measured *i*, *z*, and *y* band magnitude in the PS1 catalogue can be robustly assigned a probability of being a HzQ. In the following section, we apply our HzQ selection method to publicly available photometric data from PS1 in a bid to identify previously undiscovered HzQs at $z \gtrsim 5.5$.

4. Implementing the quasar selection algorithm

Having defined all the necessary components for our HzQ probability assignment method, in this section we apply it to data taken from PS1 as described in Section 2. When provided the i_{P1}, z_{P1} and y_{P1} magnitudes of any source, our method described above should yield a posterior probability, P_q, that the source is a HzQ.

4.1. Candidate HzQ samples

As the sky density priors give a much greater weight to non-quasar likelihoods, we must define a posterior threshold that is capable of capturing the quasar population while largely rejecting other foreground contaminants. Using PS1 photometry of known quasars at $z > 5.5$ from Bañados et al. (2016), we find that a posterior threshold of $P_q > 5 \times 10^{-4}$ accepts ~90% of the quasar population. The same threshold also rejects more than 99% of dwarf stars and low-redshift galaxies, as is shown in Figure 3.

We introduce an additional requirement ensuring a good-quality detection in PS1 to remove spurious detections, saturated
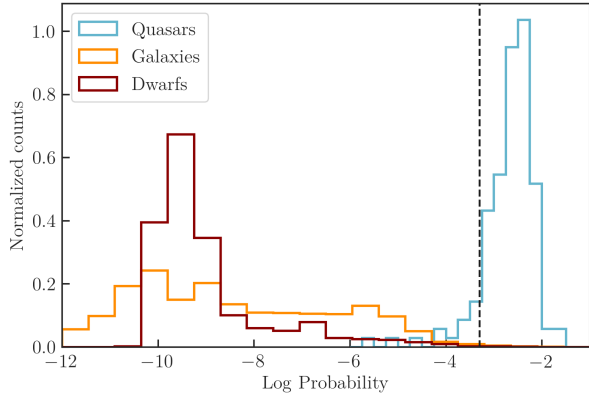
Fig. 3. Posterior quasar probabilities of the different populations assigned by our HzQ selection method. The chosen posterior threshold of $P_q > 5 \times 10^{-4}$ is shown using the dashed line, which retains $\sim 90\%$ of the known quasar population while rejecting $> 99\%$ of foreground contaminant populations.

Fig. 4. Cumulative distribution function of z_{P1} magnitudes of the sources classified as candidate HzQs (red line) compared with the full PS1 photometric sample (black line). Our algorithm preferentially assigns higher HzQ priorities to sources that are brighter and securely detected in the z_{P1} band, ensuring that the Lyα break resulting in higher $i - z$ colours is securely constrained.

counts, and other instances resulting in bad photometry in the catalogue by ensuring that the parameter `iQfPerfect` > 0.85. We exclude sources that do not adequately fit any of the modelled populations by removing sources that have low likelihood scores from all GMMs,

$$\log \sum_k \mathcal{L}_k > -10 \, . \tag{15}$$

Having established the adequate posterior threshold that maximises the chances of identifying HzQs and minimises the incidence of foreground contaminants, we now proceed to run our novel HzQ selection method on the photometric data in the *i*, *z*, and *y* bands queried from PS1. Our initial data set contained $\sim 5 \times 10^5$ sources, out of which 508 sources were selected with probability above the set threshold. Finally, 263 sources satisfied the additional good-quality detection requirement, of which 12 sources had an accompanying LOFAR detection.

To investigate the selection function introduced by our algorithm to identify candidate HzQs, in Figure 4 we show the cumulative distribution function (CDF) of z_{P1} magnitudes of sources lying above the probability threshold of being HzQs (red line), along with the CDF of the z_{P1} magnitudes of all the sources that were passed through the algorithm (black line). Very clearly, our algorithm preferentially classifies objects with brighter z_{P1} magnitudes as candidate HzQs, placing a stronger emphasis on capturing the Lyα break which manifests itself as higher $i - z$ photometric colours. The brighter z_{P1} magnitudes also ensure that sources classified as candidate HzQs are securely detected at redder wavelengths. The z_{P1} mag distribution for sources with high HzQ probabilities peaks at $z_{P1} \approx 19$, with objects fainter than $z_{P1} > 21$ very rarely selected, as it would be impossible to constrain the Lyα break in objects with the faintest z_{P1} magnitudes. The comparison shown in Figure 4 therefore serves as a validation for our new HzQ selection algorithm.

The 263 candidate HzQs lying above the posterior threshold of $P_q > 5 \times 10^{-4}$ selected from our method represent a very small fraction ($\sim 0.05\%$) of the initial data set. Our strict threshold clearly results in a drastic reduction in the number of candidate HzQs, which can subsequently be visually inspected, whereby additional spurious, extended, or otherwise undesired sources can then be rejected. The main aim of visual inspection
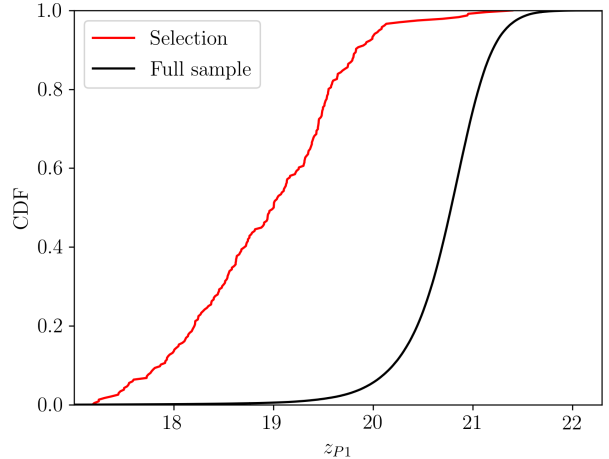
is to identify clearly spurious sources that may have been missed as such by the quality selection parameter (`iQfPerfect`). Examples include contamination by bright stellar spikes, cosmic ray residuals, and grouped bright pixels. We additionally rejected candidates that showed extended morphologies, as HzQs are highly likely to appear as point sources in PS1 images. The visual inspection was carried out by JDW, AS, and KJD, with mutual agreement being required in order to reject a candidate.

As a result, our conservative approach to visual inspection resulted in a large fraction of HzQ candidates being rejected, with 65 sources —11 of which have a radio detection— remaining as good HzQ candidates suitable for spectroscopic follow-up. The entire sample is summarised in Figure 5, where the discriminating power of the posterior calculation can be appreciated. Further details of the sources can be found in Table A.1. Noteworthy is a cluster of sources around $z - y = 1.5$, which represents a subset of the sources that have a detection in LoTSS. In total, 4417 sources in the full catalogue have a LOFAR counterpart, none of which would be selected if no radio counterpart was present. The addition of radio data has given higher significance to these sources, and shows that the method can robustly take into account the additional information provided by a radio detection.

4.2. Comparison with colour selection

In this section we test the efficacy of our Bayesian HzQ selection method compared to the traditional colour- and magnitude-based selection. We note that $\sim 30\%$ of sources that were assigned high HzQ probabilities lie below the traditional $i - z > 2$ colour cut (Figure 5), and may potentially be missed by studies relying on binary colour and magnitude cuts for HzQ searches. To compare results, we apply a colour cut of $i - z > 2$ on the full PS1 sample, which selects 634 sources. This shows that even though sources below the colour cut can be above the probability threshold, many sources above the cut are also rejected. To investigate if these sources are rightfully rejected or accepted, we can run
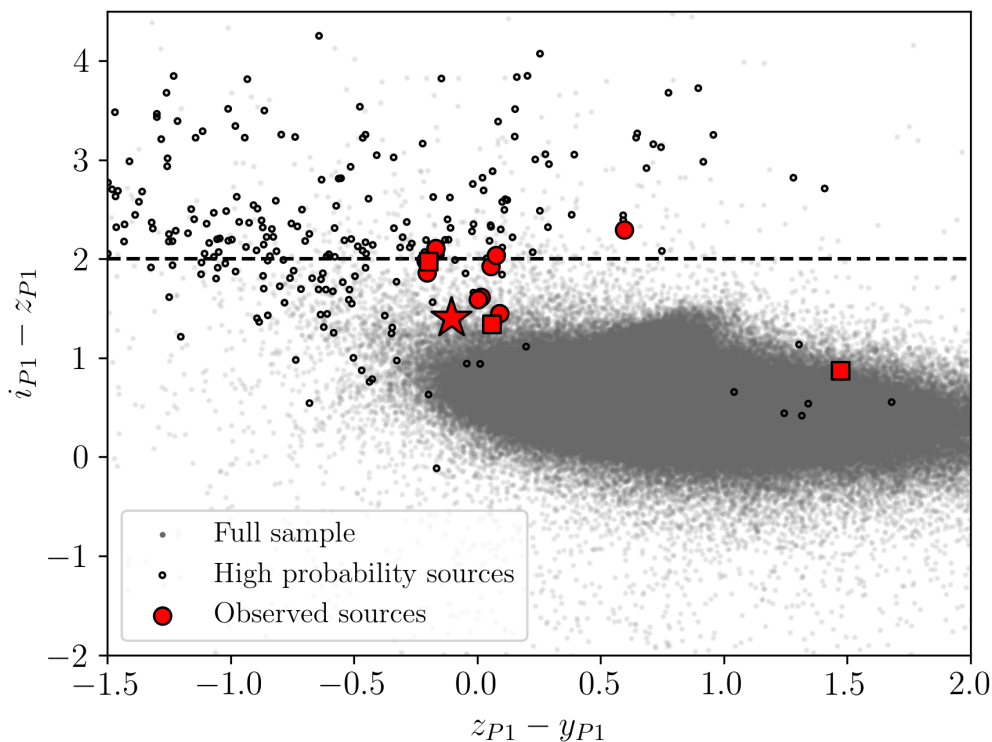
Fig. 5. Distribution of sources in the $z_{P1} - y_{P1}$ vs $i_{P1} - z_{P1}$ diagram of the full PS1 sample (grey points), high-probability HzQs (black circles), and the candidate HzQs selected for spectroscopic follow-up (red circles) and those with a radio detection (red squares). The traditional colour cut at $i - z = 2$ is marked using the black dashed line, demonstrating that our method assigns a high HzQ probability to several sources that lie below this selection. We also mark P144+50 (red star) in the colour–colour plot, which was discovered using our method at a redshift of $z = 5.66$, and with $i - z = 1.4$ lies below the standard photometric selection employed in other studies.

the algorithm on the data sets described in Section 3.1, and compare them with a $i - z > 2$ colour cut.

For the HzQ sample, we simulate 10^4 HzQs using the same method as described in Section 3.1. We also assign z-band magnitudes to the simulated quasars following a log-normal distribution based on the magnitude distribution of HzQs from Ross & Cross (2020). From the z-band magnitude, the i- and y-band magnitudes are automatically assigned based on the colours of the quasars. The algorithm is run on this sample, and as for the PS1 sample, only sources above the probability threshold of $P_q = 5 \times 10^{-4}$ are accepted. Through this we find 7614 (76%) HzQs above the probability threshold, while 7388 (74%) are above the $i - z > 2$ colour cut. Much like in the sample described above, the probability cut rejects sources above the colour cut and vice versa, such that an important difference between the methods is in which parts of colour space are probed.

As both Bayesian selection and colour selection methods return roughly the same number of HzQ candidates, we test the efficacy of selection by repeating the same experiment with the contaminant populations. For the brown dwarf population, we use the Best et al. (2017) catalogue containing ~10^4 sources, and using the above mentioned probability threshold selection, our algorithm shortlists 14 (0.15%) known brown dwarfs as candidate HzQs. However, the colour selection selects 68 (0.72%) brown dwarfs as HzQ candidates.

Using the low- and intermediate-redshift galaxy catalogues from Williams et al. (2018) containing ~10^4 sources, our algorithm classifies only one (0.01%) galaxy as a candidate HzQ,

compared to 36 (0.38%) galaxies being classified as HzQ candidates based on colour selection. Overall, we find that our method rejects a larger percentage of contaminants, while retaining a similar fraction of HzQs compared to a simple $i - z > 2$ colour selection, implying a higher overall efficacy.

We note that in $i - z$ colour-based selections, often an additional colour criterion of $z - y < 0.5$ is applied (e.g. Bañados et al. 2014) to further remove contaminants. Applying these $i - z > 2$ and $z - y < 0.5$ cuts on our simulated HzQ, brown dwarf, and galaxy catalogues, we find that all brown dwarfs are eliminated and 28 galaxies are classified as candidate HzQs. However, these cuts only retain 58% of HzQs from our simulated sample, clearly showing that although a large fraction of contaminants are eliminated using colour cuts based on i, z, and y band photometry, several HzQs may also be missed by such a selection.

The possible addition of radio data further improves the efficacy of HzQ selection, and we consider a sample where all sources have counterparts in the radio. Using the assumed radio detection rates in Section 3.4, we see almost all quasars with a radio detection are accepted (91%). This fraction is purely considering the number of sources that are above the probability threshold, which is in addition to the increase in probability for these quasars across the board. All dwarf stars are eliminated from the sample, while as before 0.01% of galaxies are retained. This shows that the addition of radio data can be extremely valuable for identifying HzQs, and significantly increases the purity of the resulting candidate sample.

In Figure 6, the results of the colour and probability selection on the test samples of HzQs (left) brown dwarfs (middle), and galaxies (right) are summarised. Here it is clear that the probability selection method handily rejects contaminants that occupy the same colour space as HzQs and would normally be included in colour selection. We note that, although both methods recover a similar number of HzQs, different subsets are selected, as the probability selection recovers a significant fraction of HzQs below the colour cut, while also rejecting a portion of HzQs above the colour cut that lie close to the colour space of brown dwarfs. As HzQs with $i - z < 2$ have generally lower redshifts, this demonstrates that the probability selection can be especially effective in recovering HzQs at $z \sim 5.6$.

5. Spectroscopic follow-up

To demonstrate the efficacy of our quasar selection method and confirm the nature of the candidate sources, we obtained spectroscopy primarily targeting the Lyα lines for the most promising HzQs identified by our selection method. Nevertheless, our final sample of high-quality HzQ candidates still contains an impractical number of sources for additional spectroscopic observations, and therefore we assigned priorities to sources in our final sample based on available independent ancillary data primarily at infrared wavelengths, which were crucially not used in the probability assignment using our method.

We first cross-matched our candidate HzQ sample with the AllWISE Data Release catalogue[3], which builds upon the data collected by the *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) mission by additionally including data from the NEOWISE surveys (Mainzer et al. 2011). The AllWISE data contain photometry in the *WISE* $W1$, $W2$, $W3$, and $W4$ bands, offering wavelength coverage in the range $3.4 - 22 \mu m$. We additionally cross-match our sample with the UKIRT Hemisphere Survey (UHS, Dye et al. 2018) data release containing deep J-band imaging and source catalogues over $\sim 12700 \deg^2$ of the sky.

We used this ancillary data to assign priorities to sources in our candidate HzQ sample for spectroscopic follow-up. First, we assigned higher priority to sources with lower $|z_{P1} - y_{P1}|$ and $|y_{P1} - J|$, which brings us closer to the locus of colours from known HzQs at $z > 5.5$. Next, based on the infrared colours seen in the sample of known HzQs from Bañados et al. (2016), which was also used as a validation sample for this study, we assigned higher priorities to sources that satisfied the following conditions:

$$W2 - W3 > 0$$
$$-0.2 < W1 - W2 < 0.85$$
$$-0.7 < y_{P1} - W1 < 2.2.$$

Lastly, the brightest sources in our sample were assigned higher priority, purely to make the process of spectroscopic follow-up more efficient. Having assigned a priority for spectroscopic follow-up to each candidate HzQ, we now describe our spectroscopic observations below.

5.1. Description of observations

The spectroscopic observations of our candidate HzQs presented in this work were obtained using the Intermediate Dispersion Spectrograph[4] (IDS) on the 2.5m Isaac Newton Telescope (INT, PI: Wagenveld, Program: N17). The observations were taken over a period of 6 nights in Spring 2019, during which 13 of the highest priority candidate HzQs were observed. Three nights were unfortunately lost due to bad weather, and the remaining three nights had favourable conditions with an average seeing of 0.5" from 6 to 8 April 2019.

The observations were taken using the R400R grating in the Red arm of the spectrograph, with a slit width of 1.5 arcseconds and slit length of 3 arcminutes. Standard afternoon calibrations were performed with both lamp and sky flats taken before each observing night. A flux standard was observed at the beginning and the end of each night. We used CuAr+CuNe lamps for wavelength calibration, which were observed at the position of each target before the sky exposure.

The targets were observed in blocks of 1800s, with total integration times per source ranging from 3600s to 7200s. Due to telescope limitations and higher priority assigned to brighter sources, only sources brighter than a z-magnitude of 20.5 were observed. Blind offsetting was used to acquire faint targets and standard data-reduction procedures, which includes bias subtraction, flat-fielding, sky subtraction, wavelength calibration, and flux calibration, were performed using a custom PYTHON-based data-reduction pipeline written by our team[5], which is based on CCDPROC (Craig et al. 2021).

Of the 13 targets observed in this run, 11 could not be conclusively classified based on the spectra obtained. In most cases, only very faint continuum was spotted with potential narrow lines. Unfortunately, the S/N of the continuum or the emission lines for these sources was not sufficient to unambiguously determine redshifts or classify the sources as either dwarf stars in the Milky Way or low-redshift galaxies. Three of the observed sources had an accompanying radio detection, but did not contain strong emission line features in their spectra. However, the clear absence of a strong Lyα line or a break blueward of Lyα in their spectra indicated that these sources were unlikely to be quasars at $z \gtrsim 5.5$.

However, high-S/N spectra were obtained for two sources in our sample, one of which was conclusively classified as a brown dwarf star owing to clear, broad absorption features in the continuum arising from molecules such as TiO (e.g. Reiners et al. 2007), which often mimic the Lyα break found in the spectra of HzQs.

The other source, PSO J144128.715+502239.463 (P144+50 hereafter), was convincingly classified as a previously undiscovered, luminous quasar at a redshift of $z = 5.66$, and in the following section we describe the observed properties of this newly discovered HzQ.

5.2. P144+50: a luminous quasar at $z \approx 5.7$

The most luminous and promising source of the candidate sample, P144+50, has a very clear point-source-like structure across the available broad band photometry, as illustrated in Figure 7. This luminous quasar was most likely missed by earlier searches owing to its relatively low $i_{P1} - z_{P1}$ colour of 1.4, which may be excluded by traditional binary colour cuts. Although no radio counterpart for this quasar was identified within the LoTSS DR1 catalogues, P144+50 was still assigned a probability of $P_q = 0.01$ from our method, demonstrating that our novel HzQ selection method is capable of assigning realistically high prob-

---

3 https://wise2.ipac.caltech.edu/docs/release/allwise/

4 http://www.ing.iac.es/astronomy/instruments/ids/
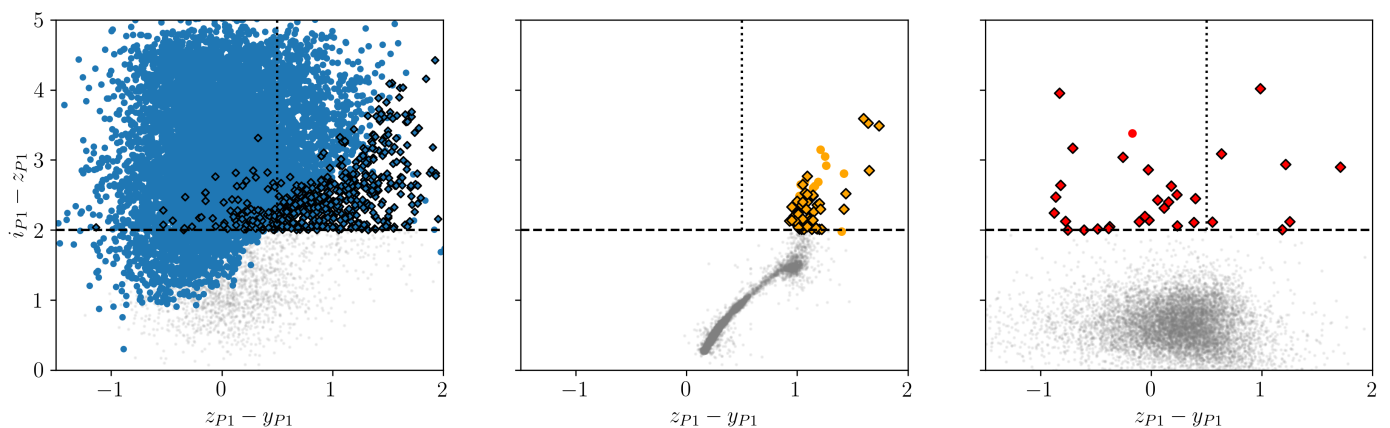5 https://github.com/aayush3009/INT-IDS-DataReduction

Fig. 6. Distribution of sources in the $z_{P1} - y_{P1}$ vs $i_{P1} - z_{P1}$ diagram of the test samples of HzQs (left), brown dwarfs (middle), and galaxies (right). Sources not selected with any method are marked in grey. Sources selected with $i - z > 2$ colour cut (dashed line) but rejected by the probability selection are marked as a diamond shape with a black edge. Notably, the $z - y < 0.5$ colour cut (dotted line) rejects all brown dwarfs from the test sample, but also a large portion of the simulated quasar sample.
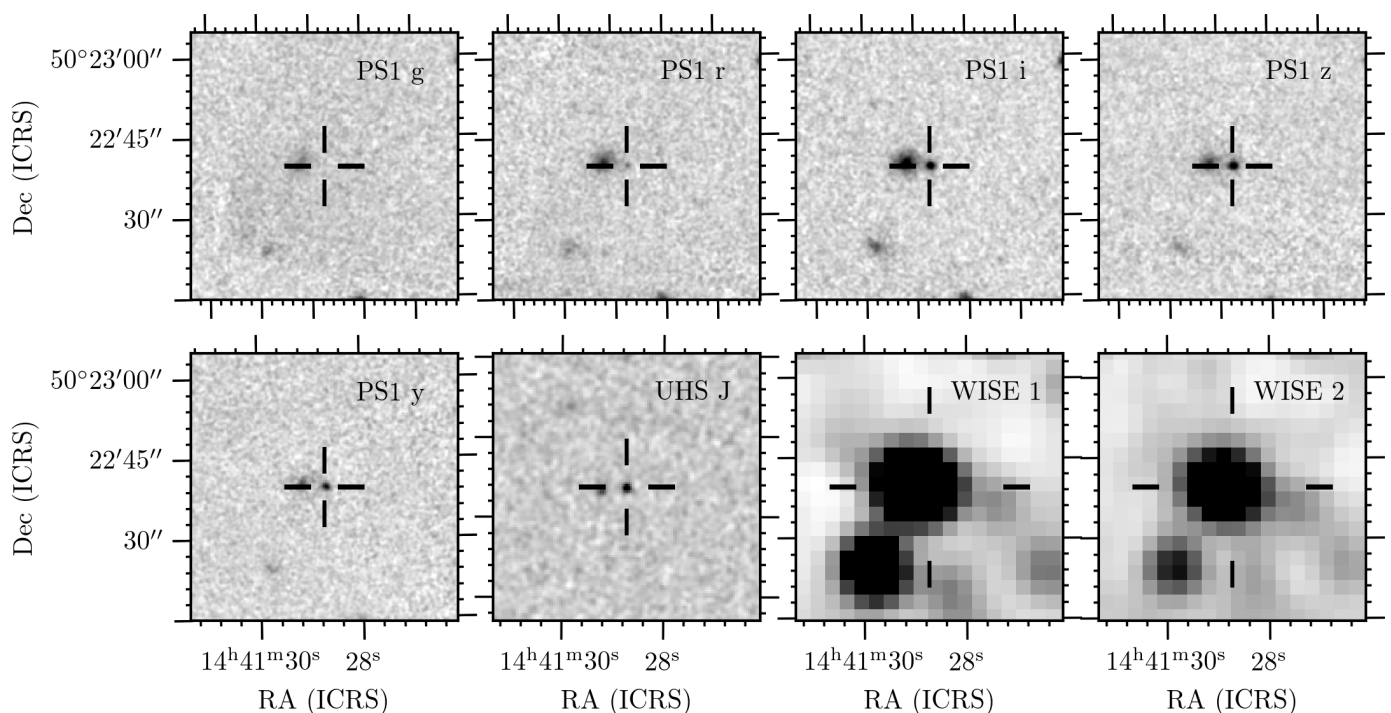
Fig. 7. P144+50 in optical and infrared bands. In the WISE images, it is blended with the neighbouring galaxy. Given the fact that the quasar is brighter in those bands while the galaxy drops off, and that its WISE magnitudes are consistent with those of other HzQs, it is likely that most of the emission in the WISE bands comes from the quasar.

abilities even to non-radio-detected HzQs. In Table 2 we give the apparent magnitudes of P144+50 in the available optical and infrared filters.

The 1D spectrum of P144+50 shown in Figure 8 displays a bright and broad, strong Lyα feature, with a clear break in the spectrum blueward of the line, showing the Gunn-Peterson trough. The peak of the Lyα line suggests a redshift of $z = 5.66$. No other rest-frame UV emission features are identified, but the Si II absorption feature may be present.

We additionally detect a Lyα forest, and faint signs of the presence of an ionised proximity zone around this QSO. Additional flux is detected around the rest-frame Lyβ and O IV wavelengths. Unfortunately, due to the limited S/N of our INT observations, any meaningful constraints on either the proximity

zone or the neutrality of the intervening IGM along the line of sight cannot be derived. Therefore, deeper follow-up observations with larger telescope facilities are required to draw robust conclusions.

Demonstrating the merits of the new HzQ selection method introduced in this paper perfectly, P144+50 is a bright hitherto undiscovered quasar at $z = 5.66$. Its rest-frame UV magnitude, $M_{1450} = -27.22$, puts it at the brighter end of the quasar luminosity function at $z \sim 6$ (Manti et al. 2017) and amongst the most luminous quasars currently known at this redshift. Given that the sky area covered by our search is ~1% of the full sky, and the sensitivity of spectroscopic observations restricting the follow-up to only those sources with magnitudes brighter than 20.5 AB, it is not improbable that other such undiscovered quasars ex-
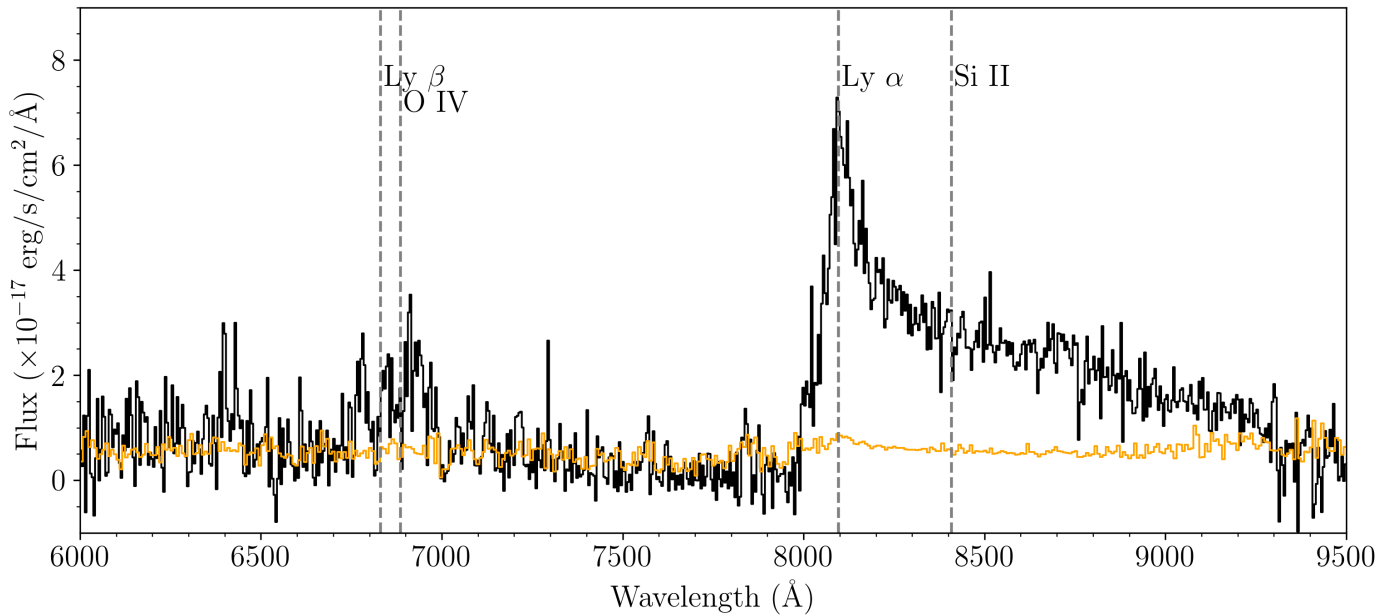
Fig. 8. 1D spectrum of P144+50 taken using the IDS instrument on the 2.5m INT. The characteristic break blueward of the Lyα line is clearly visible, unambiguously identifying it as a high-redshift quasar with a redshift of $z = 5.66$ based on its Lyα emission. Some flux is seen around the expected Lyβ and O IV features. Apart from Lyα, there are no clear emission lines visible in the spectrum due to the limited sensitivity of the INT. The orange line indicates the noise level of the spectrum.

Table 2. Observed optical and infrared magnitudes of P144+50 in all filters relevant to probability calculation and priority assignment.

Filter	Magnitude (AB)
PS1 g	> 23.2
PS1 r	> 23.2
PS1 i	20.71 ± 0.03
PS1 z	19.31 ± 0.02
PS1 y	19.41 ± 0.03
UHS J	19.34 ± 0.06
WISE $W1$	18.22 ± 0.04
WISE $W2$	18.21 ± 0.05
WISE $W3$	17.46 ± 0.28

ist within the large sky surveys that may have been missed in searches relying on binary colour selection.

Additionally, the P144+50 lies within the redshift range investigated by Yang et al. (2017), demonstrating that our GMM-based HzQ selection approach is able to successfully identify quasars within the so-called 'redshift gap' often encountered by HzQ searches employing optical broadband selection. Thanks to the increased discriminatory power provided by our algorithm, GMM-based HzQ searches might provide a powerful method for more complete samples of HzQs, including those that lie within the redshift gaps in ground-based optical broadband searches.

6. Future prospects

The Bayesian quasar-selection method presented in this work is built from priors informed by empirical relations and likelihood models from machine learning, as opposed to binary cuts in optical/infrared magnitudes and colours. As a result, this method relies heavily on the accuracy of priors derived for both HzQ populations and common contaminants in HzQ searches. Therefore, the priors can be improved in an iterative fashion by folding in the results from the ever increasing spectroscopic confirmation of candidate HzQs.

The inclusion of additional photometric data can also help to improve the priors, resulting in a more accurate HzQ probability assignment. For example, J band and *WISE* photometry for known HzQs and contaminants can be used to improve the estimation of priors, which for this work have only been used to shortlist candidate HzQs for spectroscopic follow-up. Expanding the model to include these additional dimensions can further increase its precision and reliability thanks to the increased colour information available, as well as extend its application to searches for HzQs at even higher redshifts.

The range of redshifts selected for this analysis ($5.6 < z < 6.5$) was chosen in order to enable validation with known HzQs that were selected using $i - z$ photometric colours. However, given that samples of HzQs can be simulated for training and existing HzQs can be used as validation for our selection method, other redshift ranges can be probed. For example, searches for HzQs at $z > 6.5$ can be readily carried out using our algorithm by using a handful of known HzQs at these redshifts, selected based on their $z - y$ colours. The efficacy of the method may however not be as high for selecting $z > 6.5$ HzQ candidates owing to the lower number of HzQs known at these redshifts that could be used for training and validation. As mentioned earlier, the inclusion of more photometric data may help to improve the priors for HzQs at the highest redshifts.

As the method should be extendable to larger datasets, one potential bottleneck will be excising the remaining unwanted sources after assigning probability. For the sample described in this paper, this final step was done through visual inspection of 263 high-probability candidates. For a much larger initial sample size, this method is no longer feasible. From the visual inspection we performed, most of these unwanted sources were rejected on the grounds of either being spurious, having incorrect magnitude, or appearing extended. Spurious sources are essentially removed if we force all sources to have a counter-

part in WISE and/or UHS. As these magnitudes are used anyway to assign priority to sources (Section 5), it will be doubly advantageous to implement such a selection. Incorrect magnitudes can be corrected by performing photometry directly on the Pan-STARRS images. Finally, there is clear need to differentiate between point sources and extended sources. There are several ways to do this with the Pan-STARRS catalogues, such as comparing the PSFs and Kron magnitudes of sources[6] (Farrow et al. 2013). However, these methods of differentiating between point sources and extended sources become less reliable towards lower magnitudes, where we expect more quasars. We note here that efforts are currently underway to use machine learning techniques to morphologically classify radio sources (e.g. Mostert et al. 2021), which could suitably be extended to morphological classification of candidate HzQs from optical images.

Here, we demonstrate the discriminatory power of our HzQ selection method and show that it is possible to shortlist manageable numbers of high-quality HzQ candidates from large photometric data sets. With the aforementioned flexibility and room for improvement, our algorithm can potentially be applied to even larger and deeper surveys of the sky enabled by existing state-of-the-art and upcoming ground- and space-based optical and infrared observatories such as the Vera C. Rubin Observatory (Ivezić et al. 2019), *Euclid* (Laureijs et al. 2011), the *Nancy Grace Roman Space Telescope* (formerly known as *WFIRST*; Spergel et al. 2015), and existing large surveys such as the Kilo-Degree Survey (KiDS; De Jong et al. 2013) and Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) to name a few.

Finally, while the $z = 5.66$ quasar discovered in this analysis is undetected in LoTSS radio continuum imaging, the high detection fraction of known $z > 5$ sources within the 5700 deg^2 of the forthcoming LoTSS Data Release 2 (36% at $> 2\sigma$ significance; Gloudemans et al. 2021) illustrates that the radio continuum observations can provide valuable additional information for HzQ selection and remain a powerful tool to crucially exclude contamination from Galactic dwarf stars. Relatively shallow but large-area existing radio surveys such as FIRST (Becker et al. 1995) and NVSS (Condon et al. 1998) carried out with the Very Large Array (VLA) have led to the discovery of several radio-loud quasars at $z \gtrsim 5$ (e.g. Bañados et al. 2015), and TGSS Alternative Data Release (Intema et al. 2017) covering ~ 37000 sq/ deg. of the sky at 150 MHz has already led to the discovery of the most distant radio-selected galaxy currently known (Saxena et al. 2018). The full LoTSS data release will offer sensitive radio coverage over very large sky areas in the northern hemisphere, enabling the inclusion of radio priors for a large number of candidate HzQs. These sky areas and sensitivities will be improved by upcoming ultra-deep radio surveys such as those by the Square Kilometre Array (SKA; Dewdney et al. 2009) and its precursors like MeerKAT (Jonas 2016) and ASKAP (Hotan et al. 2021) enabling even fainter radio detections.

Therefore, the HzQ selection method presented in this work is flexible, and has room for improvement given the availability of deep photometric data over large parts of the sky via existing and future large area sky surveys across wavelengths. Our method also provides a robust framework within which the additional radio information can be incorporated to potentially identify even radio-faint quasars in the early Universe.

---

[6] https://outerspace.stsci.edu/display/PANSTARRS/How+to+separate+stars+and+galaxies

7. Summary and Conclusions

In this paper we present a novel method for selecting candidate high-redshift quasars (HzQs; $z \gtrsim 5$) from large photometric data sets, making use of informed priors and Gaussian mixture models (GMMs) within a Bayesian framework. Our method attempts to capture the HzQ population more completely compared to traditionally used binary cuts in optical magnitudes and colours, while minimising the likelihood of contamination from foreground sources such as dwarf stars in the Milky Way and lower redshift dusty galaxies.

Our novel selection method builds upon previous works employing Bayesian selection of HzQ candidates using informed priors. The novelty of our methods lies in using GMMs to obtain likelihoods in optical colour–colour spaces using photometry of populations of known and simulated HzQs, as well as common contaminants such as M, L, and T brown dwarf stars and low-redshift dusty galaxies that often mimic the observed optical photometric properties of HzQs. Additional priors based on the security of optical detections, respective sky densities of the source populations, and a radio detection are used to calculate the probability of a particular source detected in a large photometric sky survey being a candidate HzQ.

We run our GMM-based HzQ search method on photometric data from the publicly available Pan-STARRS DR1 (PS1) over a limited area on the sky, coinciding with deep radio imaging from LoTSS in the HETDEX Spring field covering ~ 400 square degrees. Using in particular photometry in the PS1 i, z, and y bands, we assign candidate HzQ probabilities to $\sim 5\times10^5$ sources from PS1. Adopting a HzQ posterior probability threshold that results in the selection of $\sim 90\%$ of known HzQs at $z \gtrsim 5.5$ and the rejection of $\gtrsim 99\%$ of known foreground contaminants such as dwarf stars or low-redshift galaxies, we shortlist 263 candidate HzQs with high probabilities. By visually inspecting these candidates to spot any obvious artefacts, we select 63 sources in the final high-probability candidate HzQ sample, which can subsequently be followed up spectroscopically. To test the efficacy of the method, we run the probability selection on test samples of simulated HzQs and previously used samples of dwarfs an galaxies. We find that the efficacy of the probability method is higher than traditional colour cuts, decreasing the fraction of accepted contaminants by 86% while retaining a similar fraction of HzQs. While more stringent colour cuts decrease the contaminant fraction to levels similar to that of the probability selection, less HzQs are recovered. The efficacy of the probability selection is increased further once radio data are taken into account, reducing the fraction of contaminants by 99% compared to the traditional colour cut at the cost of selecting only quasars that have a radio detection.

Follow-up spectroscopic observations were then carried out for the highest priority HzQ candidates from our sample, with 13 candidates targeted with the 2.5m Isaac Newton Telescope. Although the nature of 11 of these 13 candidates could not be confirmed owing to low S/Ns in the relatively shallow spectra, a lack of strong Lyα emission or Lyα absorption present in the spectrum was used to rule out a very high-redshift nature.

However, the exact nature of two candidates could be established, with one being a brown dwarf star and the other being a previously undiscovered, luminous quasar at $z = 5.66$ (P144+50). The spectrum of P144+50 shows a strong and broad Lyα line, with a strong break in the spectrum bluewards of Lyα indicative of a Gunn-Peterson trough. P144+50 has a rest-frame UV magnitude of $M_{1450} = -27.22$, putting it at the very bright end of the luminosity function at this redshift. This HzQ was

likely missed by earlier searches owing to its $i - z$ photometric colour of 1.4, which falls below the traditional limits requiring $i - z > 1.5$.

The discovery of this previously undiscovered, luminous quasar at $z = 5.66$ serves as a validation of our novel HzQ selection method, indicating that a probabilistic method of selecting HzQs from large photometric surveys may perform better at returning more complete samples of HzQs as opposed to binary selections based on cuts in optical and infrared colours or magnitudes. Our method could be improved by the inclusion of more photometric data when calculating posterior probabilities, and as such can be employed on larger incoming sky surveys to discover new quasars, into the epoch of reionisation.

References

Abazajian, K., Adelman-McCarthy, J. K., Ageros, M. A., et al. 2003, AJ, 126, 2081

Bailer-Jones, C. A. L., Smith, K. W., Tiede, C., Sordo, R., & Vallenari, A. 2008, MNRAS, 391, 1838

Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, ApJS, 227, 11

Bañados, E., Venemans, B. P., Morganson, E., et al. 2014, Proceedings of the International Astronomical Union, 9, 19

Bañados, E., Venemans, B. P., Morganson, E., et al. 2015, ApJ, 804, 118

Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, MNRAS, 447, 3402

Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559

Best, W. M. J., Magnier, E. A., Liu, M. C., et al. 2017, ApJS, 234, 1

Bovy, J., Hennawi, J. F., Hogg, D. W., et al. 2011, ApJ, 729, 141

Burningham, B., Hardcastle, M., Nichols, J. D., et al. 2016, MNRAS, 463, 2202

Caballero, J. A., Burgasser, A. J., & Klement, R. 2008, A&A, 488, 181

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560

Chen, B., Stoughton, C., Smith, J. A., et al. 2001, ApJ, 553, 184

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693

Craig, M., Crawford, S., Seifert, M., et al. 2021, Astropy/Ccdproc: 2.2.0 – Image Combination Performance, Zenodo

Cristiani, S., Serrano, L. M., Fontanot, F., Vanzella, E., & Monaco, P. 2016, MNRAS, 462, 2478

De Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25

Dempster, A., Laird, N., & Rubin, D. 1977, Royals Statistical Society, B 39, 1

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEE Proceedings, 97, 1482

Duncan, K. J., Sabater, J., Röttgering, H. J., et al. 2019, A&A, 622

Dye, S., Lawrence, A., Read, M. A., et al. 2018, MNRAS, 473, 5113

Fan, X. 2006, New A Rev., 50, 665

Fan, X., Narayanan, V. K., Lupton, R. H., et al. 2001, AJ, 122, 2833

Farrow, D. J., Cole, S., Metcalfe, N., et al. 2013, MNRAS, 437, 748

Gloudemans, A. J., Duncan, K. J., Röttgering, H. J. A., et al. 2021, A&A, 656, A137

Gunn, J. E. & Peterson, B. A. 1965, ApJ, 142, 1633

Gurkan, G., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622

Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, arXiv e-prints [arXiv:astro-ph/0210394]

Holoien, T. W.-S., Marshall, P. J., & Wechsler, R. H. 2017, AJ, 153, 249

Hooper, E. J., Impey, C. D., Foltz, C. B., & Hewett, P. C. 1995, ApJ, 445, 62

Hotan, A. W., Bunton, J. D., Chippendale, A. P., et al. 2021, PASA, 38, e009

Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Jiang, L., McGreer, I. D., Fan, X., et al. 2016, ApJ, 833, 222

Jonas, J. L. 2016, Proceedings of Science, 25

Kimball, A. E., Knapp, G. R., Ivezić, E., et al. 2009, ApJ, 701, 535

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011 [arXiv:1110.3193]

Madau, P. 1995, ApJ, 441, 18

Mainzer, A., Bauer, J., Grav, T., et al. 2011, ApJ, 731, 53

Manti, S., Gallerani, S., Ferrara, A., Greig, B., & Feruglio, C. 2017, MNRAS, 466, 1160

Matsuoka, Y., Iwasawa, K., Onoue, M., et al. 2018, ApJS, 237, 5

Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2016, ApJ, 828, 26

Metcalfe, N., Farrow, D. J., Cole, S., et al. 2013, MNRAS, 435, 1825

Mortlock, D. J., Patel, M., Warren, S. J., et al. 2012, MNRAS, 419, 390

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616

Mostert, R. I. J., Duncan, K. J., Röttgering, H. J. A., et al. 2021, A&A, 645, A89

Oke, J. B. & Gunn, J. E. 1983, ApJ, 266, 713

Pipien, S., Cuby, J. G., Basa, S., et al. 2018, A&A, 617, A127

Planck Collaboration. 2015, A&A, 594, A13

Reed, S. L., Banerji, M., Becker, G. D., et al. 2019, MNRAS, 487, 1874

Reiners, A., Homeier, D., Hauschildt, P. H., & Allard, F. 2007, A&A, 473, 245

Reynolds, D. 2009, in Encyclopedia of Biometrics, Vol. 31 (Boston, MA: Springer US), 659–663

Richards, G. T., Deo, R. P., Lacy, M., et al. 2009, AJ, 137, 3884

Ross, N. P. & Cross, N. J. 2020, MNRAS, 494, 789

Saxena, A., Marinello, M., Overzier, R. A., et al. 2018, MNRAS, 480, 2733

Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, A&A, 598, A104

Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622

Silverman, B. W. 2017, Density Estimation for Statistics and Data Analysis (Boca Raton: Routledge)

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757 [astro-ph] [arXiv:1503.03757]

Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

Temple, M. J., Hewett, P. C., & Banerji, M. 2021, MNRAS, 508, 737

The Dark Energy Survey Collaboration. 2005, arXiv:astro-ph/0510346 [arXiv:astro-ph/0510346]

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549

Vedantham, H. K., Callingham, J. R., Shimwell, T. W., et al. 2020, ApJ, 903, L33

Williams, W. L., Calistro Rivera, G., Best, P. N., et al. 2018, MNRAS, 475, 3429

Williams, W. L., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622

Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, AJ, 140, 546

Willott, C. J., Delfosse, X., Forveille, T., Delorme, P., & Gwyn, S. D. J. 2005, ApJ, 633, 630

Wit, E., van den Heuvel, E., & Romeijn, J.-W. 2012, Statistica Neerlandica, 66, 217

Wright, E. L., Eisenhardt, P. R., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Yang, J., Fan, X., Wu, X.-B., et al. 2017, AJ, 153, 184

Appendix A: HzQ candidates

Table A.1. High probability HzQ candidates

Name	i_{P1}	z_{P1}	y_{P1}	LoTSS Flux (mJy)	P_q	Observed
PSO J151528.1+421313.8	21.49 ± 0.11	19.58 ± 0.04	20.03 ± 0.18		1.6×10^{-3}	
PSO J115421.7+421840.7	21.68 ± 0.12	19.32 ± 0.06	20.64 ± 0.17		5.8×10^{-1}	
PSO J150748.8+422307.8	21.50 ± 0.06	19.45 ± 0.00	19.90 ± 0.26		3.0×10^{-3}	
PSO J123718.4+422839.6	22.27 ± 0.04	20.09 ± 0.18	20.63 ± 0.16		1.4×10^{-3}	
PSO J124208.8+423946.4	22.15 ± 0.01	20.04 ± 0.15	20.21 ± 0.13		8.4×10^{-4}	Yes
PSO J124911.0+425105.3	22.11 ± 0.05	20.13 ± 0.15	20.98 ± 0.21		8.3×10^{-4}	
PSO J125047.9+430833.7	22.17 ± 0.01	19.55 ± 0.09	19.66 ± 0.14		1.3×10^{-1}	
PSO J123203.0+432745.0	21.94 ± 0.14	19.64 ± 0.17	20.47 ± 0.15		4.4×10^{-3}	
PSO J124221.7+434033.2	21.70 ± 0.06	20.27 ± 0.00	20.65 ± 0.11		1.4×10^{-3}	
PSO J122900.5+441359.8	21.61 ± 0.05	19.39 ± 0.16	20.24 ± 0.17		8.8×10^{-3}	
PSO J121800.4+453150.9	21.78 ± 0.17	19.38 ± 0.10	20.53 ± 0.15		5.2×10^{-2}	
PSO J120837.0+454149.4	22.15 ± 0.17	19.79 ± 0.16	20.55 ± 0.17		2.9×10^{-3}	
PSO J114519.9+454428.0	21.63 ± 0.04	19.93 ± 0.00	20.80 ± 0.17		2.1×10^{-2}	
PSO J112111.5+461150.9	21.47 ± 0.17	19.55 ± 0.11	20.87 ± 0.19		1.5×10^{-3}	
PSO J142738.5+473727.4	21.24 ± 0.11	20.70 ± 0.02	21.38 ± 0.00		1.1×10^{-2}	
PSO J150321.1+480022.9	21.84 ± 0.07	17.61 ± 0.01	20.50 ± 0.19		1.0	
PSO J151021.5+490023.1	21.69 ± 0.27	18.21 ± 0.02	20.70 ± 0.18		1.0	
PSO J144128.7+502239.4	20.71 ± 0.03	19.31 ± 0.02	19.41 ± 0.03		1.1×10^{-2}	Yes
PSO J112418.7+504151.3	21.78 ± 0.05	20.08 ± 0.00	20.73 ± 0.18		4.7×10^{-3}	
PSO J152639.5+520303.0	21.44 ± 0.02	17.29 ± 0.01	19.90 ± 0.16		1.0	
PSO J144047.0+520934.6	21.00 ± 0.10	18.95 ± 0.04	20.44 ± 0.17		8.3×10^{-1}	
PSO J121906.9+524229.8	21.35 ± 0.03	19.90 ± 0.02	19.81 ± 0.06		5.2×10^{-4}	Yes
PSO J120853.9+540651.1	21.37 ± 0.05	19.92 ± 0.00	20.53 ± 0.19		8.2×10^{-4}	
PSO J110945.2+574348.4	21.62 ± 0.04	17.88 ± 0.01	19.73 ± 0.17		1.0	
PSO J112328.2+595614.9	21.42 ± 0.08	18.70 ± 0.00	20.52 ± 0.18		1.0	
PSO J135335.3+600430.6	21.51 ± 0.04	20.73 ± 0.04	21.69 ± 0.00		2.5×10^{-3}	
PSO J152721.9+610352.3	21.08 ± 0.14	18.79 ± 0.01	18.19 ± 0.01		7.3×10^{-3}	Yes
PSO J141715.5+615224.3	22.00 ± 0.16	19.52 ± 0.15	20.54 ± 0.17		1.7×10^{-2}	
PSO J112052.2+472605.0	21.36 ± 0.07	19.99 ± 0.12	20.71 ± 0.16	1.10 ± 0.10	3.4×10^{-3}	
PSO J141837.2+474852.2	22.65 ± 0.35	21.32 ± 0.11	20.43 ± 0.19	3.29 ± 0.11	7.2×10^{-4}	
PSO J113104.0+475003.9	21.64 ± 0.17	19.67 ± 0.12	19.86 ± 0.15	0.44 ± 0.10	2.8×10^{-2}	Yes
PSO J123823.6+475933.1	21.13 ± 0.05	19.91 ± 0.18	20.41 ± 0.15	0.69 ± 0.15	2.1×10^{-3}	
PSO J131244.6+495724.5	21.62 ± 0.22	20.28 ± 0.16	20.23 ± 0.19	0.43 ± 0.09	7.8×10^{-4}	Yes
PSO J123626.6+501036.9	21.51 ± 0.08	20.16 ± 0.20	20.97 ± 0.41	0.64 ± 0.09	2.3×10^{-3}	
PSO J124654.9+501623.7	21.71 ± 0.19	20.65 ± 0.13	19.35 ± 0.00	0.67 ± 0.11	1.2×10^{-3}	
PSO J112037.6+502404.9	21.82 ± 0.20	20.81 ± 0.16	19.44 ± 0.22	0.45 ± 0.12	6.7×10^{-4}	
PSO J134157.7+512952.2	21.96 ± 0.19	20.93 ± 0.12	19.49 ± 0.16	31.89 ± 0.16	7.8×10^{-4}	
PSO J130926.4+525922.1	21.82 ± 0.20	20.70 ± 0.17	19.50 ± 0.17	0.97 ± 0.10	8.0×10^{-4}	
PSO J113311.2+420443.2	21.50 ± 0.04	19.68 ± 0.00	20.17 ± 0.11		1.4×10^{-3}	
PSO J123740.1+420851.0	22.06 ± 0.14	20.13 ± 0.19	20.63 ± 0.16		7.3×10^{-4}	
PSO J151948.4+423446.7	21.99 ± 0.17	19.96 ± 0.00	19.88 ± 0.18		2.1×10^{-3}	Yes
PSO J124059.8+431019.5	22.14 ± 0.18	20.26 ± 0.17	20.74 ± 0.17		7.0×10^{-4}	
PSO J140022.9+433822.2	22.11 ± 0.30	19.56 ± 0.02	18.40 ± 0.02		1.2×10^{-3}	
PSO J145612.6+442417.2	21.71 ± 0.01	20.12 ± 0.20	20.12 ± 0.10		5.4×10^{-4}	Yes
PSO J114416.9+443451.0	21.36 ± 0.05	19.50 ± 0.15	19.70 ± 0.15		2.5×10^{-3}	Yes
PSO J135622.5+453320.5	21.83 ± 0.18	19.88 ± 0.00	19.99 ± 0.18		1.5×10^{-3}	
PSO J123757.3+465507.2	21.73 ± 0.20	20.25 ± 0.04	20.61 ± 0.16		5.7×10^{-4}	
PSO J124016.5+473737.9	21.58 ± 0.03	19.55 ± 0.10	20.39 ± 0.24		1.7×10^{-3}	
PSO J124300.0+481418.3	22.81 ± 0.28	20.96 ± 0.29	20.47 ± 0.17		7.4×10^{-4}	
PSO J124203.1+495354.1	21.44 ± 0.05	19.91 ± 0.20	20.62 ± 0.18		6.3×10^{-4}	
PSO J124656.0+503223.3	21.47 ± 0.04	19.96 ± 0.17	20.63 ± 0.16		6.2×10^{-4}	
PSO J131523.9+513827.7	21.57 ± 0.07	18.58 ± 0.02	19.99 ± 0.14		2.7×10^{-3}	
PSO J143229.1+534741.1	21.94 ± 0.18	20.18 ± 0.00	20.68 ± 0.20		7.4×10^{-4}	
PSO J124450.2+585817.8	21.37 ± 0.09	19.73 ± 0.13	20.33 ± 0.08		9.4×10^{-4}	
PSO J144448.1+600520.3	21.46 ± 0.08	20.04 ± 0.17	20.46 ± 0.20		6.1×10^{-4}	

Table A.1. continued

Name	i_{P1}	z_{P1}	y_{P1}	LoTSS Flux (mJy)	P_q	Observed
PSO J131842.1+600706.5	21.67 ± 0.08	19.67 ± 0.00	20.45 ± 0.19		1.7×10^{-3}	
PSO J133347.7+603212.6	21.62 ± 0.08	20.01 ± 0.33	19.99 ± 0.04		6.1×10^{-4}	Yes
PSO J132006.7+605705.4	21.92 ± 0.18	19.81 ± 0.13	20.12 ± 0.14		2.6×10^{-3}	
PSO J131356.0+614833.6	21.65 ± 0.11	20.12 ± 0.17	20.33 ± 0.17		5.7×10^{-4}	
PSO J120732.9+492944.0	21.50 ± 0.07	20.63 ± 0.42	19.16 ± 0.05	1.16 ± 0.10	5.6×10^{-4}	Yes
PSO J115605.5+444105.3	21.68 ± 0.05	19.33 ± 0.21	20.76 ± 0.18		8.2×10^{-4}	
PSO J105545.9+445655.8	21.69 ± 0.11	19.77 ± 0.26	19.71 ± 0.06		2.6×10^{-3}	Yes
PSO J130519.1+464845.5	21.86 ± 0.20	20.22 ± 0.18	20.59 ± 0.17		5.7×10^{-4}	
PSO J124059.4+483522.9	21.70 ± 0.18	19.71 ± 0.23	19.92 ± 0.18		3.0×10^{-3}	Yes
PSO J150531.3+610408.5	21.74 ± 0.10	20.02 ± 0.13	19.70 ± 0.20		7.5×10^{-4}	