



HAL
open science

The LOFAR two metre sky survey data release 2: probabilistic spectral source classifications and faint radio source demographics

A. B. Drake, D. J. B. Smith, M. J. Hardcastle, P. N. Best, R. Kondapally, M.
I. Arnaudova, S. Das, S. Shenoy, K. J. Duncan, H. J. A. Röttgering, et al.

► To cite this version:

A. B. Drake, D. J. B. Smith, M. J. Hardcastle, P. N. Best, R. Kondapally, et al.. The LOFAR two metre sky survey data release 2: probabilistic spectral source classifications and faint radio source demographics. *Monthly Notices of the Royal Astronomical Society*, 2024, 534, pp.1107-1126. 10.1093/mnras/stae2117 . insu-04821653

HAL Id: insu-04821653

<https://insu.hal.science/insu-04821653v1>

Submitted on 6 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The LOFAR two metre sky survey data release 2: probabilistic spectral source classifications and faint radio source demographics

A. B. Drake,¹★ D. J. B. Smith,¹ M. J. Hardcastle¹, P. N. Best,² R. Kondapally,² M. I. Arnaudova¹, S. Das¹, S. Shenoy¹, K. J. Duncan², H. J. A. Röttgering³ and C. Tasse^{4,5}

¹Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK

²Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

³Leiden Observatory, Leiden University, NL-2300 RA Leiden, Netherlands

⁴GEPI & ORN, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, F-92190 Meudon, France

⁵Centre for Radio Astronomy Techniques and Technologies, Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa

Accepted 2024 August 9. Received 2024 August 9; in original form 2024 July 12

ABSTRACT

We present an analysis of 152 355 radio sources identified in the second data release of the LOFAR Two Metre Sky Survey (LoTSS-DR2) with Sloan Digital Sky Survey (SDSS) spectroscopic redshifts in the range $0.00 < z < 0.57$. Using Monte Carlo simulations, we determine the reliability of each source exhibiting an excess in radio luminosity relative to that predicted from their H α emission, and, for a subset of 124 023 sources we combine this measurement with a full BPT analysis. Using these two independent diagnostics, we determine the reliability of each source hosting a supermassive black hole of high or low Eddington-scaled accretion rate, and combine the measurements to determine the reliability of sources belonging to each of four physical classes of objects: star-forming galaxies (SFGs), radio-quiet active galactic nuclei (RQAGN), and high- or low-excitation radio galaxies (HERGs or emission-line LERGs). The result is a catalogue that enables user-defined samples of radio sources with a reliability threshold suited to their science goal, for example prioritizing purity or completeness. Here, we select high-confidence samples of radio sources (>90 per cent reliability) to report: 38 588 radio-excess AGNs in the LoTSS-DR2 sample (362 HERGs, and 12 648 emission-line LERGs), together with 38 729 SFGs, and 18 726 RQAGN. We validate these results through comparison to literature using independent emission-line measurements, and to widely adopted *WISE* photometric selection techniques. While our use of SDSS spectroscopy limits our current analysis to ~ 4 per cent of the LoTSS-DR2 catalogue, our method is directly applicable to data from the forthcoming WEAVE-LOFAR survey, which will obtain over a million spectra of 144 MHz selected sources.

Key words: catalogues – galaxies: active – galaxies: evolution – galaxies: star formation – radio continuum: galaxies.

1 INTRODUCTION

Understanding the evolution of galaxies and active galactic nuclei (AGNs), together with their interplay across cosmic time, remains a central focus for world-class facilities. Huge progress has been made in tracing the integrated star formation history (e.g. Hopkins & Beacom 2006; Madau & Dickinson 2014), now extending to the highest redshifts ($z > 6$; e.g. Kistler et al. 2009; Robertson & Ellis 2012; Fudamoto et al. 2021; Zavala et al. 2021; Bouwens et al. 2022; Algera et al. 2023; Barrufet et al. 2023), and in identifying large samples of AGNs (e.g. Mazzucchelli et al. 2017; Farina et al. 2022), including some of the first supermassive black holes in the Universe (Bañados et al. 2018; Yang et al. 2020; Wang et al. 2021). The relative importance, however, of the processes of star formation and the growth of supermassive black holes (i.e. via accretion) as function of redshift remains an open question, as indeed are

the best approaches to disentangle the signatures of each process within individual galaxies. The presence of dust inside galaxies, which is a by product of the star formation process, significantly hinders our ability to trace both star formation and AGN activity, by attenuating the light sufficiently to completely obscure these processes at optical wavelengths in many galaxies. Indeed, some studies now suggest that a large fraction of the ‘activity’ (i.e. star formation or accretion) of the Universe occurs in this ‘obscured’ phase (e.g. star formation: Magnelli et al. 2011, 2013; Gruppioni et al. 2013; Bouwens et al. 2016; Dunlop et al. 2017; Whitaker et al. 2017; Zavala et al. 2021; and AGN activity: Martínez-Sansigre et al. 2005; Perna et al. 2015; Hickox & Alexander 2018; Hatcher et al. 2021). For this reason, the radio waveband, which is impervious to dust attenuation, presents a particularly attractive solution to select samples of active extragalactic sources irrespective of their dust (or gas) content (Condon 1992). In particular, at the lowest frequencies (\lesssim a few tens of GHz) emission from active sources is primarily in the form of synchrotron radiation – either dominated by cosmic rays

* E-mail: a.drake4@herts.ac.uk

accelerated by jets, or by supernova remnants following recent star formation (at higher and lower radio luminosities, respectively).

Assessing the so-called population mix of sources as a function of radio luminosity (i.e. the relative numbers of SFGs, AGNs, and subclasses of AGNs) has been an active area of research over the past decade, as the abilities of radio observatories have rapidly evolved, leading to larger and larger samples of radio-selected extragalactic sources (e.g. Best & Heckman 2012; Shimwell et al. 2017, 2019, 2022; Smolčić et al. 2017; Sabater et al. 2019).

The demographics of the radio source population are a strong function of flux density; while initial surveys of the brightest sources were dominated by radio-loud AGNs (e.g. Laing, Riley & Longair 1983) with the latest surveys, we are now in the regime where the source population is dominated by radio-quiet AGNs and star-forming galaxies (e.g. Sabater et al. 2019; Best et al. 2023; Das et al. 2024). Essential to our ability to identify each of these physical classes, is the breadth and quality of complementary multiwavelength data sets that enable us to identify the optical/IR counterparts of the sources, determine their physical properties, and ultimately determine the dominant powering mechanism of each source. The gold standard technique is the use of optical spectroscopy to identify precise redshifts, and in the case of high-S/N spectra, determine unambiguous source classifications thanks to the characteristic emission lines associated with star formation or AGN activity.

In a landmark work, Best & Heckman (2012) applied this approach to radio sources selected across $> 7000 \text{ deg}^2$ by either the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) Sky Survey (NVSS; Condon et al. 1998) or the Faint Images of the Radio Sky at Twenty centimetres (FIRST) survey (Becker, White & Helfand 1995). The 1.4-GHz maps reached a completeness limit of $\sim 2.5 \text{ mJy}$ (NVSS; 45 arcsec beam at FWHM), and 1 mJy beam^{-1} (FIRST; 5 arcsec beam at FWHM). By combining this catalogue with the seventh data release of the Sloan Digital Sky Survey (SDSS), the authors were able to obtain optical spectroscopic data leading to 9168 cross-matched sources. In this flux regime, samples of radio galaxies are dominated by AGNs (e.g. Magliocchetti et al. 2002) and so the authors endeavoured to remove ‘contaminant’ objects dominated by star formation from the sample following Best et al. (2005) and Kauffmann, Heckman & Best (2008), before employing a multistage approach to classify the remaining ‘radio-AGNs’ (i.e. radio-excess AGNs) into high- and low-excitation classes (hereafter high-excitation radio galaxies: HERGs and low-excitation radio galaxies: LERGs). Due to the varying signal to noise in the spectra of these sources, the authors employed different techniques depending on the information available for each source, and ultimately provided a sample of 7302 radio-excess AGNs classified as either high- or low-excitation objects. Indeed part of the legacy of this work has been in building the consensus that supermassive black hole growth occurs via fundamentally different processes depending on the rate at which matter is accreted (see e.g. Hardcastle, Evans & Croston 2007; Heckman & Best 2014; Best et al. 2023). At high accretion rates ($\gtrsim 0.01$ Eddington) accretion is thought to occur via an optically thick, geometrically thin accretion disc (e.g. Shakura & Sunyaev 1973) leading to radiatively efficient accretion for which a large fraction of energy output is in the form of photons. When accretion occurs instead at a slower rate ($\lesssim 0.01$ Eddington), the accretion flow is thought to take the form of an optically thin, geometrically thick accretion disc which is radiatively *inefficient*. These sources are thought to be powered by advection-dominated accretion flows (ADAFs; Narayan & Yi 1995), and emit energy mainly in kinetic form through radio jets extending into the surrounding interstellar

medium. The most reliable approaches to distinguish between these sources have been explored extensively in the literature, but are frequently hampered by the inhomogeneity of multiwavelength data sets available and are rarely able to account for the often significant uncertainty in a particular classification.

A subsequent major leap forward came from the VLA COSMOS 3 GHz Large Project (Smolčić et al. 2017) which provided the state of the art deep-field work reaching 2 orders of magnitude deeper than Best & Heckman (2012) over a 2.6 square degree map with a mean RMS of $2.3 \mu\text{Jy beam}^{-1}$. Using exquisite panchromatic photometric observations available in the COSMOS field, Smolčić et al. (2017) analysed the population mix of 9161 radio sources with multiwavelength counterparts. The authors used a variety of criteria to separate SFGs and AGNs, and further classify the AGNs into ‘low-to-moderate’ and ‘moderate-to-high’ luminosity, as analogues of the high- and low-excitation classes outlined in Best & Heckman (2012). They report that at flux levels below $\sim 100 \mu\text{Jy}$ at 3GHz, the SFG population begins to dominate the source counts, and use the demographic breakdown of their sources to predict that future surveys such as those carried out using the SKA will provide an efficient manner in which to select large samples of SFGs.

The SKA-pathfinder ‘MeerKAT’ for instance has recently conducted the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey, and in early re-analysis of the COSMOS field, Whittam et al. (2022) find already 54 per cent of 5223 radio sources classified in their work are likely to be SFGs, becoming the dominant population below 1.4-GHz flux densities of $\approx 150 \mu\text{Jy}$.

Radio surveys now provide a means to detect vast numbers of extragalactic sources, for instance; The International LOFAR Telescope (hereafter ‘LOFAR’; van Haarlem et al. 2013), which, operating at frequencies below 200 MHz, is sensitive to a broader range of time-scales than higher frequency work, is also able to target synchrotron radiation almost entirely uncontaminated with thermal emission. LOFAR has achieved a sensitivity and survey speed which has enabled the LOFAR Two-Metre Sky Survey (LoTSS) – aiming to provide coverage of the entire northern sky at a central frequency of 144 MHz, reaching $\sim 100 \mu\text{Jy beam}^{-1}$, with a resolution of 6 arcsec FWHM. LoTSS also includes a deep tier aiming to achieve depths of $10 - 15 \mu\text{Jy beam}^{-1}$, across 25 square degrees in total. The LOFAR Deep-Fields (ELAIS-N1, Lockman Hole and Bootes) are complemented with the deepest multiwavelength data sets available in the northern sky, enabling panchromatic SED-fitting to identify the optical counterparts to the radio sources, and infer their properties. As a consensus on the physical picture of AGN accretion modes has emerged in the literature (e.g. Best et al. 2014; Hardcastle 2018; Hardcastle & Croston 2020), it has become prudent to search for AGN signatures via different approaches designed to detect characteristics of the two different modes. Indeed the signature of these fast- and slow-accretors manifests very differently in the observable properties of sources (although also see Whittam et al. 2022 and Kondapally in preparation, for a discussion of sources potentially occupying a continuum of states between the fast and slow modes). In recent work, citeBest2023 and Das et al. (2024) apply SED-fitting codes to photometry of LOFAR-detected samples in the LoTSS deep fields, and aim to identify AGN thought to be rapidly accreting (characterized by a hot, radiatively efficient accretion disc inside a dusty torus) in the SED fits, before searching independently for the likely more slowly accreting (i.e. thought to be radiatively inefficient) AGN via an excess of radio emission relative to that predicted from their SFRs. Via this method Best et al. (2023) showed that 144-MHz flux densities (F_{144}) below 1 mJy are dominated by SFGs, comprising 90 per cent of the population at $100 \mu\text{Jy}$.

In contrast to the depth of the LoTSS Deep Fields, the wide LoTSS tier is complemented with far shallower photometry, but benefits from high-quality spectroscopic sky surveys; e.g. SDSS (York et al. 2000). At present 4 per cent of LoTSS-wide sources in the second data release (DR2; Shimwell et al. 2022) have been observed with the SDSS or BOSS spectrographs. The diagnostic power of the full LoTSS data set (> 4 million sources in the LoTSS DR2 catalogue; Hardcastle et al. 2023) will be dramatically increased as the spectroscopic information available for these sources becomes more complete. A dramatic increase in the optical spectroscopic follow-up of LoTSS DR2 will be provided by the WEAVE-LOFAR survey (Smith et al. 2016), using the William Herschel Telescope’s (WHT) new massively multiplexed optical spectrograph, the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012), which is currently in the final stages of commissioning. At the sensitivity now routinely achieved with LOFAR, the wide tier of the LoTSS surveys is sensitive to a flux range and volume likely to contain every physical class of extragalactic object (radio-loud AGNs, radio-quiet AGNs, and SFGs), making several aspects of source classification of particular importance. First, our ability to classify sources with a homogeneous method that can be applied to the full sky, secondly, providing *secure* classifications (for which the gold-standard is accompanying optical spectra), together, finally, with a measure of reliability of a particular classification. Bearing in mind the imminent and voluminous data set from WEAVE-LOFAR and from other wide-area spectroscopic surveys such as that from the Dark Energy Spectroscopic Instrument (DESI), it is important to develop tools now to cleanly identify sources with a homogeneous method which can be applied to large optical spectroscopic data sets coupled with sensitive low-frequency radio data. In addition, various science goals require the selection of samples of objects in a particular class in different ways, for instance prioritizing purity or completeness, and thus a probabilistic element to classifications (i.e. the reliability of given source belonging to particular class of objects) offers a well-defined way to examine different subsets of the population.

In this work, we develop and validate a probabilistic spectroscopic classification scheme with which to homogeneously classify large samples of radio-selected sources, and derive the source demographics of sources from the LoTSS DR2 wide tier using optical counterpart associations from Hardcastle et al. (2023) and existing SDSS spectra processed and presented in the Portsmouth catalogue (Thomas et al. 2013). The paper proceeds as follows: in Section 2, we describe the data set generated by LOFAR and the subset of objects observed with SDSS and analysed by Portsmouth, in Section 3 we present our classification scheme before validating the method in Section 4. In Section 5, we present the demographics of the faint radio source population from LOFAR, and compare to previous studies, finally in Section 6 we summarize our results. Throughout this work, we assume a Λ CDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2 DATA

2.1 LoTSS

The LOFAR Two Metre Sky Survey (LoTSS; Shimwell et al. 2017, 2019, 2022) is the flagship extragalactic survey conducted by the LOFAR International Telescope (van Haarlem et al. 2013) using the high-band antennas (HBA) at 120–168 MHz with a central frequency of 144 MHz. LoTSS aims to survey the entire northern sky reaching average sensitivities of $\approx 100 \mu\text{Jy}$ at most declinations,

with a resolution of 6 arcsec using Dutch stations alone. The second data release from LoTSS (DR2; Shimwell et al. 2022) consists of 841 LOFAR pointings, and covers 5700 deg^2 of the extragalactic sky (≈ 27 per cent) focused on areas of high Galactic latitude. DR2 is currently the largest radio catalogue in existence, containing 4.2 million sources at $\geq 5\sigma$ significance. The process of assigning optical counterparts to this catalogue is described in Hardcastle et al. (2023). In brief, the likelihood-ratio cross-matching method developed for DR1 is applied (Williams et al. 2019) following approximately the same decision tree to flag sources for which additional input via the citizen science project ‘Radio Galaxy Zoo (LOFAR)’ is required. Any deviations from the workflow described in Williams et al. (2019) are highlighted in section 4 of Hardcastle et al. (2023). At the conclusion of this process, optical counterparts are identified for 85 per cent of sources in LoTSS DR2 and 58 per cent have a good redshift estimate from existing SDSS spectroscopy, or from photometric redshifts (Duncan et al. 2021).

2.2 Existing spectra from SDSS and BOSS

Optical spectroscopic information for a subset of LoTSS sources is available from the SDSS (York et al. 2000). To enable a cross-match to the largest number of sources possible, we make use of two data releases – SDSS-DR8 (Aihara et al. 2011) and BOSS-DR12 (Eisenstein et al. 2011). Based on these catalogues, Thomas et al. (2013) provide line flux measurements using the fitting-code ‘GANDALF’ (Sarzi et al. 2017) of the major optical emission lines typically used for the identification of radiative AGNs: $\text{H } \beta_{4861}$, $[\text{O III}]_{5007}$, $\text{H } \alpha_{6563}$, and $[\text{N II}]_{6584}$ (hereafter $\text{H } \beta$, $[\text{O III}]$, $\text{H } \alpha$, and $[\text{N II}]$) together with their associated errors, and fit warnings. In addition, we make use of the value-added MPA-JHU catalogue (Brinchmann et al. 2004)¹ primarily for the purpose of examining the derived physical properties of a subset of the objects we classify, specifically their star formation rates (SFRs) and stellar masses based on the methods of Kauffmann et al. (2003) and Tremonti et al. (2004).

2.3 Sample selection

The parent sample is the catalogue of 4.2 (3.6) million radio sources in the LoTSS DR2 (cross-identified) catalogue which is described in detail by Hardcastle et al. (2023). Fig. 1 shows the distribution on sky of the LoTSS DR2 sources (black outline) in comparison to SDSS-DR8/BOSS-DR12-observed objects in the Portsmouth spectroscopic catalogue (Thomas et al. 2013, shown in blue). The dark blue area shows the intersection of LoTSS DR2 and Portsmouth catalogues. To associate the spectroscopic information with the low-frequency radio data, we perform a positional cross-match between the positions of the optical counterparts identified for the 144 MHz sources in Hardcastle et al. (2023), and those in the Thomas et al. (2013) catalogue, using a nearest-neighbour algorithm with a maximum search radius of 1 arcsec. This gives us a sample of 208 816 LOFAR sources with SDSS spectroscopic information, representing 5.7 per cent of the 3672 413 LOFAR sources that fall within the region covered by the Portsmouth SDSS DR8/DR12 catalogues.

We then apply a series of cuts before proceeding to classify sources, the numbers of sources removed at each stage are recorded in Table 1. First, as a key part of our workflow (discussed in Section 3) compares the $\text{H } \alpha$ luminosity ($L_{\text{H}\alpha}$) with the 144-MHz luminosity (L_{144}), we remove sources which fall at redshifts where $\text{H } \alpha$ is

¹https://www.sdss4.org/dr17/spectro/galaxy_mpajhu/

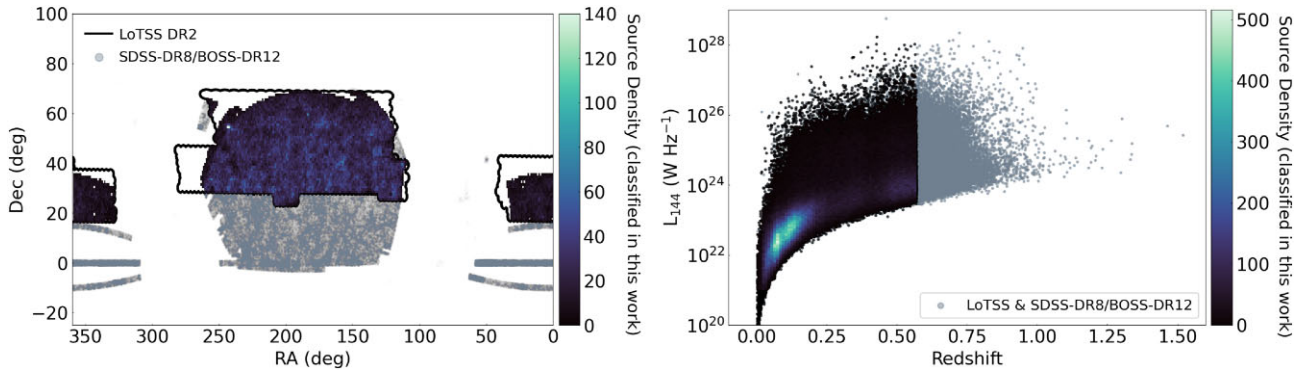


Figure 1. Left: The sky coverage of LoTSS DR2 (Shimwell et al. 2019; Hardcastle et al. 2023) given by the black outline, and the Portsmouth SDSS catalogue (Thomas et al. 2013) in greyscale. The objects studied in this work occupy the common region between these two surveys, and are shown as a density plot. Right: The redshift–luminosity plane, populated by sources common to LoTSS DR2 and the Portsmouth catalogues (in grey). Overlaid is a density plot to show the sources studied in this work. Note that we apply redshift cuts to ensure the $H\alpha$ spectral line is visible to the spectrograph before further analysis (see Section 2.3) – this is clearly visible at an upper redshift of $z \approx 0.57$. The source density in both panels is given by the colour bars to the right.

Table 1. A summary of the pre-processing steps applied to the initial matched catalogue combining Hardcastle et al. (2023) and SDSS-DR8 and BOSS-DR12 (Thomas et al. 2013). The first column describes the cut applied (as discussed in detail in Section 2.3), the central and right-hand columns list how many sources were removed in each step, and how many remain in the catalogue afterwards. In the bottom two lines, we re-iterate the final numbers of sources in the full radio-excess catalogue, and the fully classified catalogue.

Pre-processing step	Sources removed	Sources remaining
Initial match	–	208 816
Redshift cuts	34 956	173 860
f $H\alpha$ and e $H\alpha$ = 0.0	50 826	154 145
Infinite/missing $H\alpha$	101	154 044
Redshift– $H\alpha$ plane	168	153 876
R50 cut	1521	152 355
Missing BPT line	28 332	124 023
Final Radio-Excess Catalogue	–	152 355
Final Fully-Classified Catalogue	–	124 023

not visible to the SDSS and SDSS-BOSS spectrographs. For SDSS DR8², the red end of the spectrograph cuts off at 9200 \AA giving an upper redshift cut of $z = 0.385$. Likewise, for the BOSS³ subset, spectrograph transmission drops off at $10\,400\text{ \AA}$, and we impose an upper redshift cut at $z = 0.570$. We next remove any remaining sources for which the reported $H\alpha$ emission line measurement is unreliable (i.e. a flux and uncertainty equal to zero in the Portsmouth catalogue, or a reported $H\alpha$ flux that is infinite or entirely missing). Examining the redshift– $\log_{10}(L_{H\alpha})$ plane reveals several objects with unphysical Balmer line luminosities; to remove these objects, we consider only those with $4 < \log_{10}(L_{H\alpha}/L_{\odot}) < 13$.

Finally, visual inspection reveals multiple sources that are selected as high-probability radio excess but are actually nearby SFGs. This arises because the $H\alpha$ measurement is made in the 3-arcsec SDSS fibre aperture (i.e. containing only the centre of the galaxy) while the catalogue radio flux is representative of the total. To limit the influence of this effect, we consider the distribution of the effective radius, r_{50} , from the Legacy Surveys imaging as a function of redshift; and choose to remove sources with $r_{50} > 15$

arcsec. It may be possible to improve upon this with more sophisticated modelling; however, we defer that analysis for a future investigation.

The final catalogue contains 152 355 sources, within which we define one further subset – the objects for which we can apply full line-ratio analysis to determine a physical classification. To do so we exclude all objects for which any of the principal BPT diagnostic emission lines ($H\alpha$, $H\beta$, $[O\text{ III}]$, and $[N\text{ II}]$) are recorded to have a flux and associated error equal to zero in the Portsmouth catalogue (of course the requirement is already imposed on the $H\alpha$ line in the previous step). This subset contains 124 023 sources with spectral measurements in all four of the BPT lines. Each of these pre-processing steps are recorded in Table 1, and show how many sources are dropped at each stage from the initial match between the LOFAR DR2 catalogue (Hardcastle et al. 2023) and the SDSS spectra (Thomas et al. 2013).

3 CLASSIFICATION METHOD

The sample of extragalactic radio sources detected by LOFAR in the Wide LoTSS tier reaches radio flux densities which previous works have shown to be a mixture of star-forming galaxies and low-level AGN activity (e.g. Best et al. 2023; Das et al. 2024). As discussed above, coupled with the volume observed in LoTSS DR2 alone, LoTSS-Wide is sensitive to the full range of physical classes of objects. The ability to separate SFGs from AGN-dominated sources as well as distinguish between modes of AGN activity thus becomes increasingly critical. Although much has now been achieved in the LoTSS Deep-Fields using photometric data, we aim here to be able to identify the signature of an AGN in any given source via either accretion mode using optical spectroscopic data from SDSS (rather than enacting a search for a particular class of object and discounting other sources). We search first for radio excess relative to star formation activity, and then consider the emission line properties of the sources via the BPT diagram for objects exhibiting all four diagnostic lines. In addition to providing ‘maximum likelihood’ classifications, this method allows us to quantify the doubt associated with a given classification, though a Monte Carlo simulation in which we generate 1000 realizations of our matched catalogue and combine the resultant classifications of the individual catalogues into an overall probabilistic set of classifications. We describe each stage of this process in the following sections.

²<https://classic.sdss.org/dr1/products/spectra/index.php>

³<https://www.sdss.org/instruments/boss-spectrographs/>

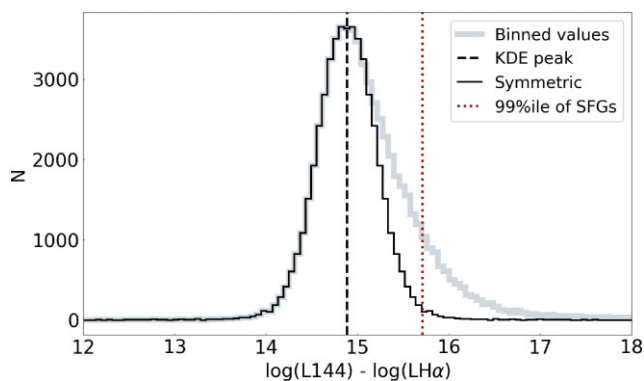


Figure 2. Histogram showing the distribution in $\log_{10} L_{144} - \log_{10} L_{H\alpha}^{\text{corr}}$ of sources with $> 5\sigma$ detections in both LoTSS and their $H\alpha$ emission line (shown in light grey). Overlaid is the KDE estimate of the modal value, indicated by the vertical dashed line. The estimated distribution for SFGs is shown in black, obtained by symmetrizing the left-hand side of the distribution about the KDE peak, and also overlaid is the 99th percentile of the inferred SFG distribution, above which we classify objects as having a 144 MHz excess (red dotted line).

3.1 Identification of sources with a radio excess

The relationship between SFR and radio luminosity arising from synchrotron emission generated in supernova remnants (shocks) is a well-studied and calibrated phenomenon (e.g. Bell 2003) including investigation with low frequency radio data (e.g. Gürkan et al. 2018, Smith et al. 2021). This relationship offers us a means through which to predict the level of radio emission associated with star formation for a particular source given a reliable SFR estimate. Sources with a radio luminosity in excess of the level predicted for their SFR can then be flagged as exhibiting a radio excess inferred to arise from nuclear activity (i.e. an AGN). The most appropriate way to quantify this excess has been investigated in the literature by several authors e.g. Sabater et al. (2019). Here, we compare the radio luminosity at 144 MHz (L_{144}) with the $H\alpha$ luminosity in solar units corrected for extinction at $H\alpha$ via the Balmer decrement (i.e. see left-hand panel of Fig. 3). To do this we calculate values of extinction, A_v , for all sources with both the $H\alpha$ and $H\beta$ lines recorded in the Portsmouth catalogue, and apply the standard Balmer correction for all physical (positive) values of A_v (in the case of a negative value of A_v , which is unphysical, no extinction correction is applied). For those objects without an $H\beta$ measurement, we instead apply a statistical extinction correction using the distribution of A_v values for all objects exhibiting a 5σ detection in both the $H\alpha$ and $H\beta$ lines. For each of our realizations, we then draw an extinction correction randomly from this distribution for each object without an $H\beta$ measurement. A similar method was also implemented by Best & Heckman (2012), and Sabater et al. (2019) on the first data release of LoTSS in the HETDEX Spring Field (Hill et al. 2008). Fig. 2 presents a histogram of the ratios of radio to $H\alpha$ luminosities of sources with $> 5\sigma$ detections in both LoTSS and the $H\alpha$ emission line (plotted as $\log_{10} L_{144} - \log_{10} L_{H\alpha}^{\text{corr}}$). By approximating the SFG population as a Gaussian distribution symmetric about the peak estimated from the kernel density distribution (i.e. mirroring the left-hand side of the distribution around the KDE peak) we can infer the 99th percentile of the SFG distribution, and define objects lying to the right of this as having a radio excess. This means that at the position of the division line 99 per cent of SFGs will lie to the left in ($\log_{10} L_{144} - \log_{10} L_{H\alpha}^{\text{corr}}$). Sources to the right of the line are assumed to exhibit a radio-excess due to nuclear activity. Identification of a

radio excess in isolation, does not have implications for the accretion rate or mode of the AGN; i.e. sources with both high- and low-Eddington-scaled accretion rates (radiatively efficient or inefficient AGN) can show a radio excess relative to their SFRs.

We define sources as having a radio excess if:

$$\log_{10}(L_{144\text{MHz}}/\text{WHz}^{-1}) > \log_{10}(L_{H\alpha}^{\text{Corrected}}/L_{\odot}) + 15.71 \quad (1)$$

3.2 Identification of radiative AGNs

The identification of radiatively efficient AGNs has been attempted and refined via various different methods depending on the data available. When spectroscopic data are available the presence of this accretion disc can be inferred via optical emission-line ratios, most frequently via the so-called ‘BPT’ diagram (Baldwin et al. 1981). Ionization by the hot material in the inner-parts of the accretion disc excites species of atoms with higher ionization energies than possible from ionizing stellar radiation alone, altering the ratios of pairs of emission lines of different species (pairs of lines are selected at similar wavelengths such that differential dust extinction along the spectrum is not an issue). The demarcation between objects powered purely by star formation, and those dominated by light from an accretion disc has been investigated thoroughly by multiple authors, with one of two divisions mainly being adopted depending on the science question at hand. In this work, we adopt the Kauffmann et al. (2003) division line (which is empirically derived to trace the upper boundary of the SFG locus in the $N[\text{II}]/H\alpha - [\text{O III}]/H\beta$ plane) to class objects as radiative AGNs (see Fig. 3 right-hand panel). We also consider the Kewley et al. (2001) maximal starburst line that uses a combination of stellar population synthesis (Fioc & Rocca-Volmerange 1997) and photoionization models (Sutherland & Dopita 1993) to derive a theoretical upper limit for galaxies purely driven by star formation. Sources with emission line ratios robustly measured to lie beyond this line definitively require the presence of a radiative AGN in order to reproduce the source’s position on the BPT diagram, whereas those lying between the two divisions are considered to be ‘composite’ sources, with spectra likely to exhibit signs of both star formation and radiative AGN activity. An additional separation can be made between objects with hard-ionising spectra characteristic of Seyfert galaxies (e.g. Weedman 1977), and those characteristic of lower ionisation radiative AGN – Low Ionization Nuclear Emission Regions (LINERs; e.g. Heckman 1980; Cid Fernandes et al. 2011). This means every source with all four BPT emission lines measured will fall into one of three classes – BPT_SFG, BPT_SEY (Seyfert), or BPT_LIN (LINER). Some fraction of BPT_SEY and BPT_LIN sources will fall in the composite region between the Kauffmann et al. (2003) and Kewley et al. (2001) demarcation lines. Thus we also report three subsets of the above categories BPT_COMP (subset of BPT_SEY and BPT_LIN which fall in the composite region), and further divide BPT_COMP into those sources on the Seyfert and LINER sides of the division line proposed in Kauffmann et al. (2003); BPT_CSEY and BPT_CLIN respectively.

3.3 Combination into overall source classification

The above diagnostics provide two independent axes along which to infer the presence of an AGN. First via an excess of radio emission beyond that expected for its SFR – a method sensitive to all radio-loud AGNs and some radio-quiet AGNs via low-level jets or accretion discs (this is the primary method used to detect radiatively inefficient i.e. jet-mode AGNs). Secondly, for those sources with the four diagnostic BPT emission lines measured in the Portsmouth

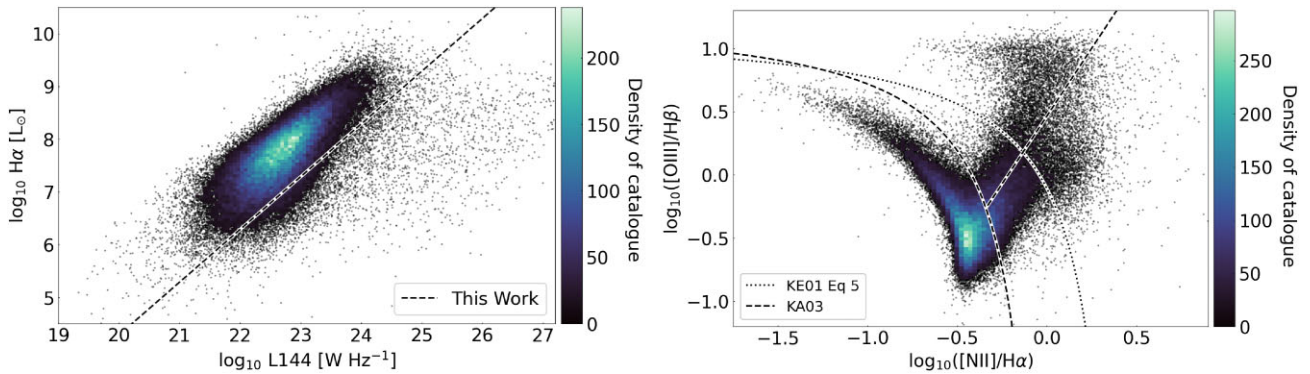


Figure 3. The two diagnostic diagrams used for our probabilistic source classification. Left: the Balmer-decrement corrected $H\alpha$ luminosity compared to the 144 MHz luminosity. This plot is used to identify those (realizations of) sources exhibiting an excess in radio luminosity above what might be expected on the basis of the $H\alpha$ luminosity (i.e. star formation in the host galaxy). Right: the BPT diagram (Baldwin, Phillips & Terlevich 1981) used to identify those (realizations of) sources which contain radiative AGN, and distinguish between high- and low-excitation sources for those objects with recorded fluxes in all four diagnostic lines (at any S/N ratio). See text for details.

catalogue, the presence of a radiative AGN source becomes apparent via its emission line ratios. The combination of these two axes then allows us to classify every emission-line source in our sample into one of the following four physical categories: (1) SFGs, (2) RQAGN (containing both Seyferts and LINERS), and two classes of radio-excess AGN – (3) high-excitation radio galaxies (HERGs), and (4) emission-line low-excitation radio galaxies (LINELERGs).⁴ Table 2 shows how the results of these two diagnostics are combined into each of the physical classes listed above. As discussed in Section 3.5, we release with this paper our full table of diagnostics, allowing the user to construct a particular physical class of objects according to their own criteria. For instance the population of LERGs can be considered in a number of ways – while some LERGs are found to be quiescent galaxies with large jets, others are seen to exhibit emission lines consistent with being LINERS, and recent work in the LoTSS Deep Fields has identified a subset of star-forming LERGs (Kondapally et al. 2022).

3.4 Probabilistic classification of the faint radio source population

An important aspect of this work is our treatment of the uncertainty associated with the classification of objects into particular classes. As all measurements used in our analysis are accompanied by uncertainties (i.e. flux errors on the radio flux at 144 MHz and on all measured emission lines in the Portsmouth catalogue), we can simply perform a Monte Carlo simulation to evaluate the range of possible outcomes. We generate 1000 realizations of the full catalogue, by perturbing the radio and line fluxes in each realization by randomly drawing samples from normal distributions with standard deviation equal to the reported uncertainties. We then apply the full classification workflow to each of the thousand catalogues individually, before combining the results to determine what fraction of realizations of each source belong to each particular class. In addition to the overall source classifications, we report the probabilistic results of each individual diagnostic – radio excess for all objects with a measured $H\alpha$ flux, and BPT line ratios for all objects with all four BPT diagnostic lines measured. In Fig. 4, we show the reliability

⁴Note that in order to be placed on the BPT diagram sources must have measured emission lines, and so we refer to this class as ‘LINELERGs’—likely to be a subset of the full LERG population (see Sections 3.5 and 5.1)

distribution of the four physical classes following the construction of our probabilistic source classifications. The reliability reported along the horizontal axis depicts the fraction of realizations that have fallen into this class, such that each object is represented once in each of the four different classes (i.e. four times in total in this plot). An object which for instance is classified as an SFG in 80 per cent of the catalogue realizations will fall into the bin at a reliability of 0.8 in the grey histogram, if for instance in 10 per cent of realizations it is instead classified as a RQ AGN it will appear at a value of 0.1 in the gold histogram, and for example twice more at 0.05 in the blue and red histograms if the remaining 10 per cent of realizations results were distributed evenly between the HERG and LERG classes. One takeaway from Fig. 4 can be seen on the right-hand side of the plot; $\sim 40\,000$ sources are securely classified (in >90 per cent of realizations) as SFGs, $\sim 20\,000$ as RQ AGN, $\sim 7\,000$ as LINELERGs, and ~ 400 as HERGs.⁵

3.5 Catalogue description

With this paper, we release our probabilistic spectral source classifications for the 152 355 sources in common between LoTSS DR2 and SDSS-DR8/BOSS-DR12 with a measured $H\alpha$ flux. In Table 2, we list the column headings of our catalogue with a short description of each in turn. We provide three different warnings to caution the user of the potential pitfalls associated with a given object’s classification. `CLASS_Z_WARNING` provides an alert that the spectroscopic redshift recorded by Portsmouth differs from the `z_best` estimate of Hardcastle et al. (2023) (used to derive radio luminosities) by $\Delta z > 0.001$. The `BALMER_CORR_WARNING` indicates that our Balmer-correction on the $H\alpha$ luminosity has been carried out statistically as a consequence of a missing $H\beta$ flux in the Portsmouth catalogue (see Section 3.1). Finally, we provide the `z_score` warning, to indicate the significance of deviation from the null hypothesis that maximum likelihood classification is correct. This warning applies to some low S/N spectra for which

⁵It is also interesting to note how the LINELERGs are divided between those with SFG-like and LINER-like emission. For objects with `CLASS_LINELERG` > 0.90 , we find that for 50 per cent of realizations or more: ≈ 30 per cent are classed as SFGs on the BPT diagram, ≈ 64 per cent are classed as LINERS, and the remainder switch between the SFG and LINER regions.

Table 2. Column headings included in the table of classifications we provide with this work. In the first column we list the column heading, in the second column a brief description of its content, and in the final column we describe the values the column can take.

Column heading	Description	Values
CLASS_z_WARNING	Redshift warning: $(z_{\text{best}} - z_{\text{Ports}}) > 0.001$	Binary 0/1 (1 = warning)
RADIO_EXCESS	Fraction of realizations in which $L_{144\text{MHz}}$ is larger than predicted values for 99 per cent of SFGs (using Balmer-corrected $L_{\text{H}\alpha}$)	Decimal between 0–1
BALMER_CORR_WARNING	Balmer correction derived without H β detection. Extinction is instead drawn randomly from the distribution of values calculated for sources with 5σ H α and H β detections.	Binary 0/1 (1 = warning)
BPT_SFG	Reliability of line ratios lying in SFG area of the BPT diagram ¹	Decimal between 0–1
BPT_SEY	Reliability of line ratios lying in Seyfert area of the BPT diagram ¹	Decimal between 0–1
BPT_LIN	Reliability of line ratios lying in LINER area of the BPT diagram ¹	Decimal between 0–1
BPT_COMP	Reliability of line ratios lying in Composite area of the BPT diagram ^{1,2}	Decimal between 0–1
BPT_CSEY	Reliability of line ratios in Composite-Seyfert area of the BPT diagram ^{1,2}	Decimal between 0–1
BPT_CLIN	Reliability of line ratios in Composite-LINER area of the BPT diagram ^{1,3}	Decimal between 0–1
BPT_ML	Maximum likelihood BPT result (i.e. using Portsmouth catalogue fluxes)	s, S, L, C, X, Y ⁴
zscore	Significance of deviation from the null hypothesis that maximum likelihood classification is correct	Decimal
CLASS_SFG	Reliability of CLASS_SFG classification – CLASS_SFG requires no radio excess to have been measured, and a BPT line ratio falling in the SFG part of the diagram	Decimal between 0 and 1
CLASS_RQAGN	Reliability of CLASS_RQAGN classification – CLASS_RQAGN requires that no radio excess is measured, and a BPT line ratio above the Kauffmann et al. (2003) demarcation line	Decimal between 0 and 1
CLASS_HERG	Reliability of CLASS_HERG classification – CLASS_HERG requires that a radio excess is measured, and that the BPT line ratio lies in the Seyfert region of the plot	Decimal between 0 and 1
CLASS_LINELERG	Reliability of CLASS_LINELERG classification – CLASS_LINELERG requires that a radio excess is measured, and that the BPT line ratio lies in either the SFG or LINER regions	Decimal between 0 and 1

Notes. ¹ Reliability calculated as fraction of catalogue realizations falling in this class.

²BPT_COMP is the sum of the subsets BPT_SEY and BPT_LIN which fall between the Kauffmann et al. (2003) and Kewley et al. (2001) demarcation lines.

³BPT_CSEY and BPT_CLIN classes occupy the composite area of the BPT diagram, and fall to the left and right of the Seyfert LINER division, respectively.

⁴ s=SFG, S=Seyfert, L=LINER, C=Composite, X=composite-Seyfert, Y = composite-LINER.

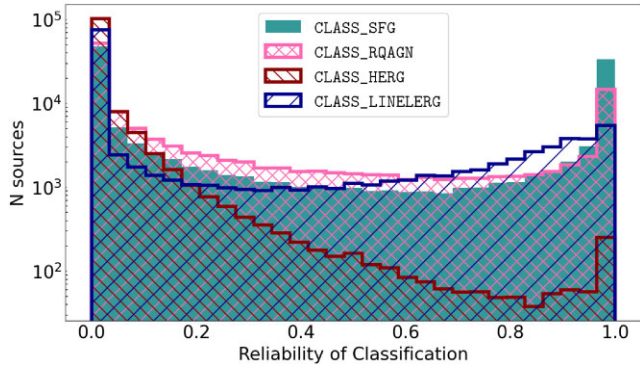


Figure 4. The distribution of CLASS_SFG, CLASS_RQAGN, CLASS_HERG, and CLASS_LINELERG for our sample, shown as teal shading, pink, red, and blue hatching, respectively, as indicated by the legend.

the nature of BPT classification (i.e. using ratios of line intensities) leads to non-intuitive behaviour in the Monte Carlo realizations of the catalogue. For instance, when the flux measurement of one diagnostic line in a ratio is close to zero, but the other line is not, the realizations of this ratio may lie far from the measured line ratios (and consequently the maximum-likelihood classification). We provide in the RADIO_EXCESS column a number between 0 and 1 indicating the number of catalogue realizations for which a radio excess was measured for each source in our catalogue. For those sources with all four of the diagnostic BPT lines recorded by Portsmouth we also provide a number between 0 and 1 in the BPT_SFG, BPT_SEY, BPT_LIN, BPT_COMP, BPT_CSEY, and BPT_CLIN columns indicating the fraction of times that each source was classified into these parts of the BPT diagram. We also include BPT_ML that records the maximum-likelihood BPT classification of each source (i.e. where the Portsmouth-measured line ratios fall). Finally, in the CLASS_SFG, CLASS_RQAGN, CLASS_LINELERG, and CLASS_HERG columns, we provide our probabilistic spectral source classifications into four physical classes. As we describe in Table 2, the classes are constructed as follows in each individual realization of the catalogue (before calculating their probabilistic ‘average’ results): CLASS_SFG requires no radio excess to have been measured, and a BPT line ratio falling in the SFG part of the diagram. CLASS_RQAGN requires that no radio excess is measured, and a BPT line ratio above the Kauffmann et al. (2003) demarcation line. CLASS_LINELERG requires that a radio excess *is* measured, and that the BPT line ratio lies in either the SFG or LINER regions. Finally to be classed as CLASS_HERG requires that a radio excess *is* measured, and that the BPT line ratio lies in the Seyfert region of the plot.

We also provide in the Appendix (Table A1) a snippet of the catalogue provided with this work, and some examples of how to construct a particular class of objects.

4 VALIDATION

In this section, we will compare the results of our classifications with other works and diagnostics.

4.1 Excitation index distribution for HERGs and LERGs

One of the key tools available for studying radio excess AGN is the Excitation Index (hereafter EI; Buttiglione et al. 2010) which is designed to measure the relative intensity of high- and low-excitation

emission lines in a source’s optical spectrum. Estimating a source’s EI requires measurements of six emission line species (H β , [O III] λ 5007, [O I] λ 6364, [N II] λ 6583, H α , and the [S II] λ 6716, 6731 doublet), combined such that:

$$\text{EI} \equiv \log_{10}([\text{OIII}]/\text{H}\beta) - \frac{1}{3}[\log_{10}([\text{NII}]/\text{H}\alpha) + \log_{10}([\text{SII}]/\text{H}\alpha) + \log_{10}([\text{OI}]/\text{H}\alpha)] \quad (2)$$

with the dividing line between high- and low-excitation sources defined to be at EI = 0.95 (Buttiglione et al. 2010). We decided against adopting an EI criterion in our classification scheme since the requirement for detections of *six* line species was likely to dramatically reduce the number of sources that we are able to classify. Nevertheless, since this is one of the key diagnostics of the ionization state of radio excess sources (e.g. in Best & Heckman 2012) it is clearly of interest to see how our results compare for the subset of sources for which we have the necessary number of detections.

The distribution of EI for all 6596 sources with the necessary six emission lines measured at $\geq 3\sigma$ is shown as the light grey histogram in Fig. 5, with the canonical dividing line between low- and high-excitation sources overlaid as the vertical dashed line. However, only 289 of these sources are classified as having a radio excess (i.e. RADIO_EXCESS > 0.90 in our catalogue), and their EI distribution is overlaid in dark grey. This represents just 0.007 per cent of the 38 588 sources for which we are able to identify a radio excess at this reliability by comparing their H α and 144 MHz luminosity.

We have also overlaid the EI distributions of sources which our classification scheme identifies as having CLASS_HERG and CLASS_LINELERG > 0.90 as red and blue hatched histograms, respectively. It is encouraging to see that the mean EI for the HERGs and LINELERGs offset from one another, but also that the canonical dividing line of EI = 0.95 falls very close to the point where the two histograms cross. Indeed, using this subset with 3σ detections in all six EI lines, we find that 84 per cent of the HERGs have EI > 0.95, while 97 per cent of the LINELERGs have EI < 0.95.

In the right-hand panel of Fig. 5, we show sources with matches in the MPA-JHU catalogue (Brinchmann et al. 2004) to demonstrate the distribution of EW[O III] values as a function of E.I. As previously, we show only those sources for which all six E.I. emission lines were measured with $> 3\sigma$ significance, but this time they must also have a match in the MPAJHU catalogue (a total of 5 931 sources). Overlaying the subsets of these sources which are also identified as CLASS_HERG or CLASS_LINELERG > 0.90, we find that the sources scatter in approximately the expected locations. We have denoted on the plot some previously suggested divisions between the HERG and LERG populations in terms of [O III] equivalent width – 3 Å (Laing et al. 1994) and 10 Å (Tadhunter et al. 1998). The lower division line does seem to cleanly select securely identified LERG-like objects, while the upper suggested division is more akin to a clean selection of HERG-like objects. Substantial mixing between securely-classified CLASS_HERG and CLASS_LINELERG objects is seen between the two lines.

Finally, in Fig. 6 we show the two diagnostics that we have used to classify our sample (the BPT diagram is shown in the left-hand panel, with the radio excess plot shown in the right-hand panel), where we have colour coded all sources that are unlikely to be SFGs (i.e. those that have BPT_SFG < 0.1) but with $\geq 3\sigma$ detections in the EI emission lines by their EI. The colour scaling has been chosen such that sources with EI > 0.95 are shown in red, and those with EI < 0.95 are in blue. It is immediately apparent that the areas of the BPT diagram that we call ‘high-excitation’ (namely the Seyfert area) are

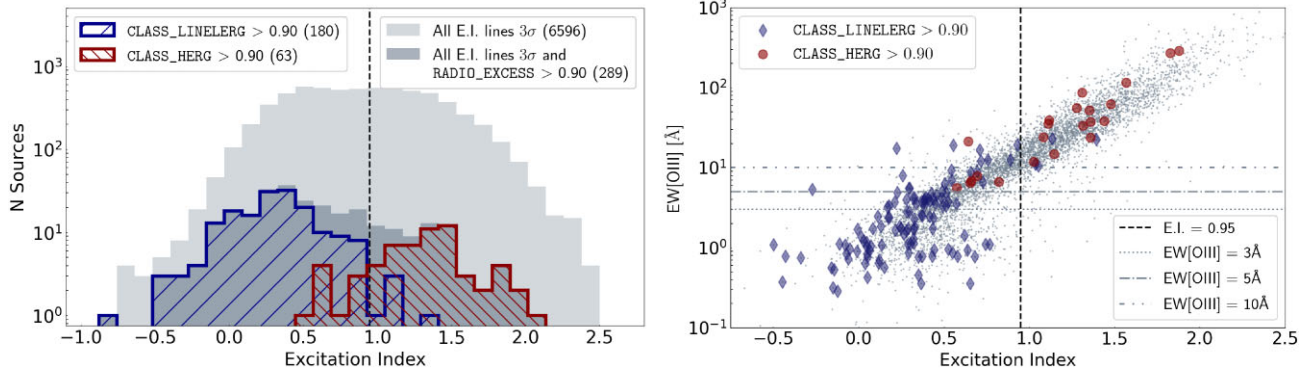


Figure 5. Left: The distribution in Excitation Index for all sources with $\geq 3\sigma$ detections in all of the necessary emission lines (see text; shown in light grey) overlaid with the subsets of sources which have $\text{RADIO_EXCESS} > 0.90$ (shown as blue and red hatched histograms, respectively). The vertical dashed line at $\text{E.I.} = 0.95$ indicates the threshold used by Buttiglione et al. (2010) to distinguish between high- and low-excitation sources. The legend indicates the number of sources in each class. Right: the distribution of sources in the $\text{E.I.}-\text{EW}[\text{O III}]$ plane. In grey, we show objects in our catalogue with a match in the MPA-JHU catalogue (Brinchmann et al. 2004), again with all E.I. emission lines at $> 3\sigma$. We overlay the positions of the subsets of these objects with CLASS_LINELERG and CLASS_HERG showing a reliability > 0.90 . Horizontal dashed lines denote the various suggested divisions in the literature Laing et al. (1994) at 3 \AA , Best & Heckman (2012) at 5 \AA , and Tadhunter et al. (1998) at 10 \AA .

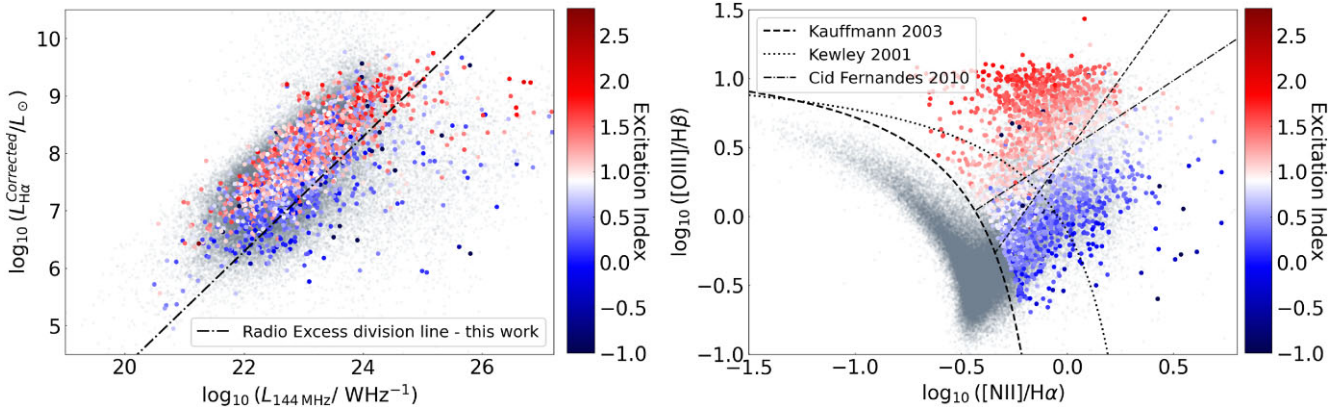


Figure 6. The location of sources in the diagnostic diagrams used for our classification (i.e. the $\text{H}\alpha$ versus 144 MHz luminosity diagram and the BPT-[N II]) shown as a function of excitation index, indicated by the colour bar to the right for sources with $\text{BPT_SFG} < 0.1$. The distribution for the full population is shown for reference in the background in light grey to give context. In the right-hand panel, the diagnostic lines from Kauffmann et al. (2003), Kewley et al. (2001), and Cid Fernandes et al. (2010) are shown as the dashed, dotted, and dotted–dashed lines, respectively. The dotted–dashed line in the left-hand panel represents the dividing line above which we identify a source as having a radio excess, given in equation (1).

dominated by sources with $\text{EI} > 0.95$, and conversely that the LINER area (lower right in the BPT diagram, which we define as containing low-excitation sources) are dominated by sources with blue colours, signifying $\text{EI} < 0.95$. This was also noted by Buttiglione et al. (2010); however, the sample used in that work contained only 113 bright sources (building on the previous works of Hine & Longair 1979; Laing et al. 1994), which meant that it was not possible to compare the two methods at the time.

Finally, we note that there is an interesting possible correlation between the position in the $\text{H}\alpha$ versus 144 MHz luminosity plot and the EI for the sources that do not show a radio excess. It appears as though the sources with lower $\text{H}\alpha$ luminosity (and therefore lower SFRs assuming a linear correlation, e.g. Kennicutt & Evans 2012) appear to have weaker high-excitation lines than the higher $\text{H}\alpha$ luminosity (more vigorously star-forming) counterparts. This is in keeping with expectations that star formation and radiative AGN activity go hand-in-hand, as noted by e.g. (Gürkan et al. 2015).

This good agreement between EI and our method offers significant encouragement that we are able to identify high-excitation sources,

even in the absence of detections in all six lines required for calculating the EI.

4.2 Comparison with Best and Heckman catalogue

An additional test of how well our method performs, is to compare our results with (Best & Heckman 2012, hereafter BH12). To do this, we first identify common sources by performing a positional cross-match between the BH12 source positions, and the positions of optical counterparts to the LoTSS sources in the Hardcastle et al. (2023) catalogue, using a nearest neighbour algorithm with maximum search radius equal to 1 arcsec. We find that 4772 sources satisfy the matching criteria (3279 of with all four BPT lines recorded in the Portsmouth catalogue enabling a full classification in this work). This includes 116 BH12 HERGs, 1864 LERGs, and 927 sources which BH12 classified as ‘star-forming galaxy’ (however we note that in B12 this classification also includes radio-quiet AGN, on the basis that they assume a stellar origin for the radio emission in those sources).

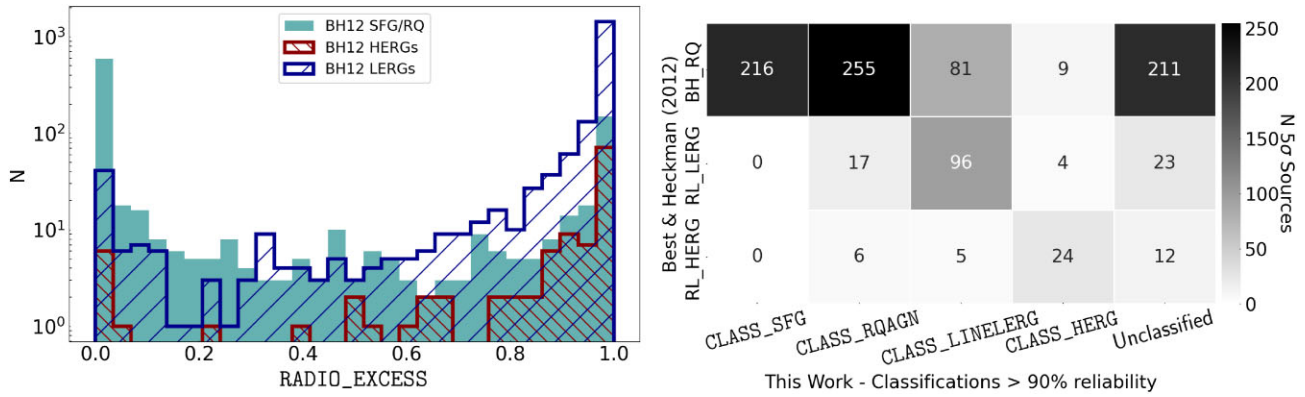


Figure 7. Left: The distribution of our `RADIO_EXCESS` parameter for sources identified as SFG/radio-quiet objects, HERGs and LERGs in Best & Heckman (2012), coloured teal, red and blue, respectively (as indicated in the legend). Right: a confusion matrix to compare final classifications of objects between BH12 and this work. The objects contained exhibit 5σ detections in all four BPT lines (in the Portsmouth catalogue), and have been classified with > 90 per cent reliability in this work, with the remainder listed in the ‘unclassified’ column. See text for details.

The distribution of our `RADIO_EXCESS` parameter for sources classified as SFG, HERG, and LERG by BH12 are shown as the solid teal, blue, and red hatched histograms in Fig. 7. It is very encouraging to note that 86 per cent of the sources that B12 classify as RLAGN are assigned `RADIO_EXCESS` values > 0.90 in our catalogue.

However, when we look at how the different BH12 source types are classified overall (i.e. including the BPT information), it doesn’t look quite as encouraging:

- (i) 55 per cent of the 927 BH12 SFGs/Radio-quiet objects have `CLASS_SFG` or `CLASS_RQAGN` > 0.90 ,
- (ii) 25 per cent of the 116 BH12 HERGs have `CLASS_HERG` > 0.90 , and
- (iii) 58 per cent of the 1864 BH12 LERGs have `CLASS_LINELERG` > 0.90 .

The BH12 classification scheme includes a complex workflow that first attempts to classify sources as high- or low-excitation on the basis of their EI, BPT line ratios (using the standard $[\text{N II}]/\text{H}\alpha$ and $[\text{O III}]/\text{H}\beta$ axes) and $\text{EW}_{[\text{O III}]}$ properties being at least 1σ away from some criterion. If a source does not meet the 1σ criterion in all of these three categories, then the steps are repeated without the 1σ criterion. This is a pragmatic approach to classifying as many sources as possible, but also means that the BH12 classifications are maximum-likelihood estimates, and often have significant uncertainties associated with them that are not visible *a posteriori*, for instance when using the catalogue. To provide a fairer comparison of the two approaches, we repeat the analysis above, this time considering only those sources which have 5σ detections in each of the four BPT lines in the Portsmouth catalogue in order to identify those sources which have the most reliable classifications in the BH12 catalogue. If we do this, we find that:

- (i) 61 per cent (67 per cent) of the 772 BH12 SFG have `CLASS_SFG` or `CLASS_RQAGN` > 0.90 (> 0.80),
- (ii) 51 per cent (60 per cent) of the 47 BH12 HERGs have `CLASS_HERG` > 0.90 (> 0.80), and
- (iii) 69 per cent (71 per cent) of the 140 BH12 LERGs have `CLASS_LINELERG` > 0.90 (> 0.80).

Taking this 5σ sample, and considering only those sources which are ‘reliably’ classified in our catalogue (reliability $> 90\%$) we construct the confusion matrix shown in the right-hand panel of Fig. 7. This gives us chance to examine the classification of sources for

which we would like to make a reliable (and correct) classification. Considering the top row first, BH12 5σ ‘SFGs’/radio-quiet objects (772 sources), it is encouraging to see that 216 and 255 of these objects are encapsulated by our `CLASS_SFG` or `CLASS_RQAGN` classes, respectively. A further 90 objects have been reliability classified into one of the radio-excess classes (81 in `CLASS_LINELERG` and 9 in `CLASS_HERG`) – as we have deliberately re-defined the threshold above which we consider objects to have a radio-excess (with a physically motivated division based on the distribution of $L_{144-L_{\text{H}\alpha}}$ values) this is not surprising. The remainder of the 772 BH12 5σ ‘SFGs’/radio-quiet objects (211 sources) evidently have reliabilities in each class < 90 per cent reflecting the uncertainty in their classification. This is the expected behaviour from our approach and quantifies the uncertainty on classifications that might be considered secure via S/N alone. In the central row, of BH12 5σ LERGs (140 sources), it is reassuring to see that 96 objects have been reliably classified into `CLASS_LINELERG`. 17 objects however have shifted to being reliably classified as `CLASS_RQAGN`, which, similarly to above, can be attributed to our updated definition of the division between radio-quiet and radio-excess objects. No objects have been classified reliably as `CLASS_SFG`. Perhaps more puzzlingly, 4 objects are reliably classified into `CLASS_HERG` in this work. These objects are not ambiguously placed on the BPT diagram, nor do they exhibit an E.I. which would class them as a HERG in BH12. It is true, however, that BH12 were using a previous version of an SDSS catalogue, and so small differences in flux measurements (or their uncertainties) may shift these objects across diagnostic boundaries or into subsequent parts of the BH12 workflow. Again the remainder of the 140 5σ BH12 LERGs (23 sources) have been classified with reliabilities < 90 per cent in any class, reflecting the ambiguity in their ultimate classification. In the bottom row of BH12 5σ HERGs (47 sources), 24 sources are reliably classified as `CLASS_HERG`, another 6 objects have shifted to `CLASS_RQAGN` due to the updated radio-excess threshold, and 5 objects have been reliably classified as `CLASS_LINELERG`. The shifts between `CLASS_HERG` and `CLASS_LINELERG` are not intuitive, and are most likely to have arisen from differences in the reported flux measurements of the two different catalogues.

Fig. 8 shows the location of this BH12 5σ sample in our diagnostic diagrams. In both plots, the full population in our catalogue with $\geq 5\sigma$ detections in all four BPT emission lines is shown in light grey, and has been overlaid with the BH12 ‘SFGs’/radio-quiet

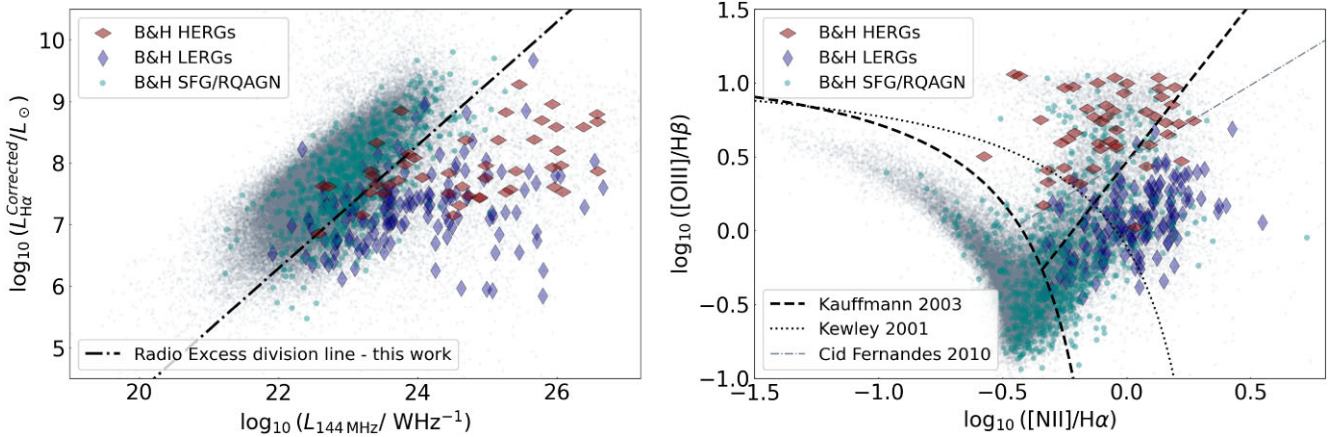


Figure 8. The $L_{H\alpha}$ versus L_{144} (left), and BPT (right) parameter spaces for sources with 5σ detections in all four BPT lines in the Portsmouth catalogues. We overlay sources identified as SFGs/radio-quiet, HERGs and LERGs in Best & Heckman (2012), coloured teal, red, and blue, respectively (as indicated in the legend). The full population of 5σ sources in our catalogue is shown as the grey distribution in the background.

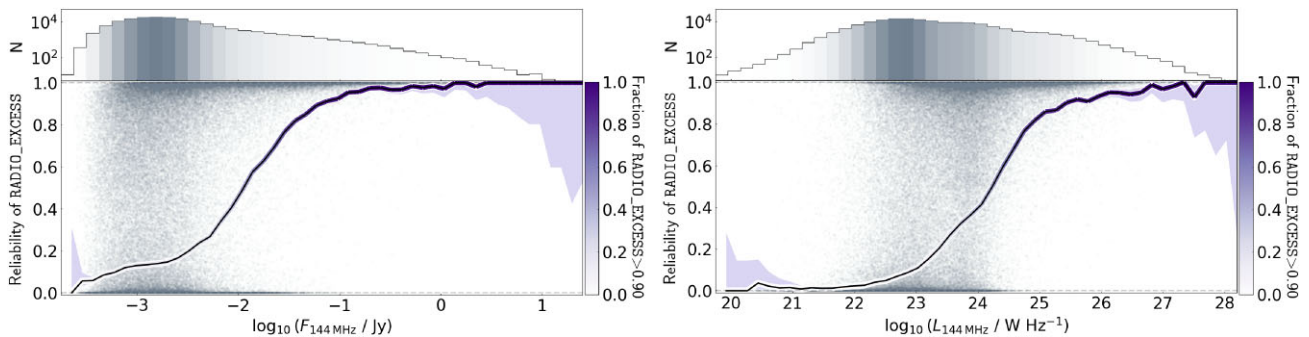


Figure 9. The variation in the RADIO_EXCESS parameter as a function of (left:) 144 MHz flux density and (right:) luminosity. In the main panels, the greyscale scatter points indicate the source positions in this plane, and these have been overlaid with the cumulative count of sources passing the 90 per cent reliability threshold in our RADIO_EXCESS parameter in bins of flux density/log luminosity. The purple shading indicates the 90 per cent confidence interval on this fraction. The upper panels indicate the overall distributions of flux density and luminosity, on the same horizontal axes. The histograms are shaded to scale linearly with the number of sources in each bin.

objects, LERGs and HERGs (shown as teal, blue, and red symbols, respectively). This is useful information since, as touched-upon above, in addition to the different workflow implemented, BH12 used the MPA-JHU catalogue from Brinchmann et al. (2004) rather than the later updated Portsmouth catalogue described by Thomas et al. (2013). It is certainly possible for line measurements to differ between the two catalogues given the significantly different methodologies, however, we see excellent general agreement, with BH12 HERGs and LERGs residing predominantly in the expected regions of the BPT diagram, and the BPT SFGs consistent with a mix of sources which are genuinely SFG (according to our definition) and those which host a radiative AGN but no radio excess.

Comparing the $H\alpha$ and radio luminosity methods between the two works is more complicated, since (aside from using a different radio frequency of 1.4 GHz from NVSS) BH12 also compare the strength of the 4000 Å break (measured using the $D_n(4000)$ parameter in the MPAJHU catalogue) with the stellar-mass-normalized 1.4 GHz luminosity (see also Sabater et al. 2019). Nevertheless, it is clear that the great majority of BH12 HERGs and LERGs lie in the radio-excess area of this diagnostic diagram (to the right of Fig. 8), while the majority of the BH12 SFGs reside in the region expected for SFGs.

4.3 Further tests of probabilistic classification

Fig. 9 shows the relationship between our RADIO_EXCESS parameter (i.e. the fraction of realizations for a given source that have a 144 MHz luminosity excess) and the 144 MHz flux density and luminosity, in the left-hand and right-hand panels, respectively. The individual source properties are shown as the greyscale points in the background, and these have been overlaid with the cumulative count of sources passing the 90 per cent reliability threshold in our RADIO_EXCESS parameter, with a 90 per cent confidence interval indicated by the purple shading. As expected, we find that the brightest and most luminous sources virtually all exhibit a radio excess, consistent with our expectations that the brightest sources should be predominantly RLAGN (e.g. Whittam et al. 2022; Best et al. 2023). Similarly, the radio-excess fraction is much smaller among the faintest and least-luminous sources. In the upper panels, we also include histograms to depict the distribution of sources in terms of their 144 MHz flux density and log luminosity on the left and right, respectively. The histogram’s shading scales linearly with the number of sources – this is particularly of use in the right-hand panel, as it becomes evident that the majority of sources reside in the $\log L_{144} \approx 23$ region, which is not necessarily clear in the scatter plot.

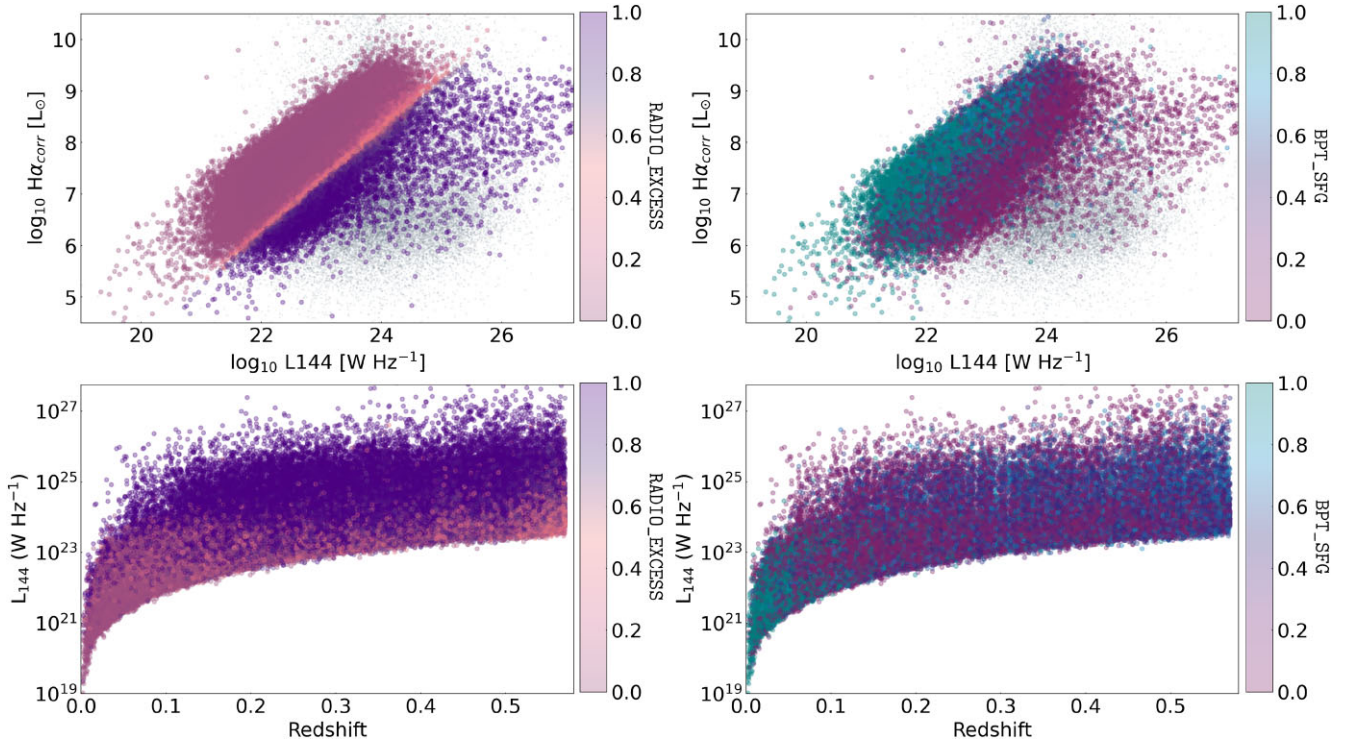


Figure 10. Top left (right) The 144 MHz– $H\alpha^{\text{corrected}}$ Luminosity plane colour coded by the RADIO_EXCESS (BPT_SFG) parameter, i.e. the fraction of realizations for which we identify a source as having a radio excess (/being identified as a radiative AGN on the BPT diagram). These two quantities represent the two independent axes along which we identify signatures of AGN activity. In the lower left (right) panel, we show the relationship between z and $L_{144\text{MHz}}$, again colour coded by RADIO_EXCESS (BPT_SFG).

In the top panels of Fig. 10, we show the comparison between $H\alpha$ and 144 MHz luminosity for all sources in our sample, which are shown as the grey background image. Overlaid are the sources with $\geq 5\sigma$ detections in both the 144 MHz LoTSS data and the $H\alpha$ emission line, and these have been colour-coded by the RADIO_EXCESS and BPT_SFG parameter in the left-hand and right-hand panels, respectively. By design, we see a larger radio excess as we move to the right side of the left-hand side plot, however, unlike previous works, our Monte Carlo simulations have enabled us to determine the reliability of a radio excess for every source accounting for the uncertainties in both quantities. In the right-hand panel, the distribution of BPT_SFG is as we would expect; towards more reliable values of BPT_SFG objects cluster in the radio-quiet portion of the plot, and at the lowest values of BPT_SFG (i.e. selecting all radiative AGN) sources dominate the plot towards the highest radio luminosities, and are intermingled with the radio-quiet portion of the plot due to the radio-quiet radiative AGN selected via the BPT diagram.

In the lower panels of Fig. 10, we show how our sample fills the redshift versus 144 MHz luminosity plane. Each point has again been colour-coded by the RADIO_EXCESS (left) and BPT_SFG (right) parameter. As expected on the basis of Fig. 9, in the left-hand panel we see that radio excess sources are more prevalent at higher 144 MHz luminosities across virtually the full redshift range and virtually absent at the lowest redshifts in our sample. An interesting feature is visible in the RADIO_EXCESS values visible for sources at the highest end of our redshift range – the reduced certainty over whether sources are radio excess could be a result of the lower SNR ratio in the $H\alpha$ fluxes that we have been able to measure for the most distant BOSS sources. In the right-hand panel, we show

the same redshift–luminosity plane; however, this time sources have been colour-coded by their BPT_SFG values (i.e. the fraction of the 1000 Monte Carlo realizations of their emission line and 144 MHz properties for which they are classified as star-forming galaxies in the standard BPT plot). As expected, the majority of the SFGs identified with the highest confidence (i.e. $\text{BPT_SFG} > 0.9$) are local and it becomes increasingly difficult to identify SFGs at high confidence above $z \approx 0.25$ (see also Section 5.1, below).

4.4 Physical properties of source by classification

A subset of our sample was included in the value-added catalogue constructed by the MPA-JHU group (Brinchmann et al. 2004). While we chose to classify our objects using the emission-line measurements presented in the Portsmouth catalogue (which superseded the MPA-JHU measurements), we can use the MPA-JHU physical parameters (e.g. SFR and stellar mass; M_*) as independent validation of our classification technique. In Fig. 11 for instance we show two panels to demonstrate the distribution of our securely classified objects (applying a 90 percent threshold) in the $\log_{10}(M_*/M_\odot)$ versus $\log_{10}\text{SFR}$, and $\log_{10}\text{SFR}$ versus $\log_{10}L_{144}$ planes. In the left-hand panel, we see that the objects with a reliable classification as CLASS_SFG form a ‘main sequence’ of SFGs in $\log(M_*/M_\odot)$ versus $\log_{10}\text{SFR}$, which aligns well with the parametrization from Schreiber et al. (2015, evaluated at $z = 0.2$ and converted to the initial mass function from Kroupa 2001, which was adopted for the MPA-JHU catalogue) which has been overlaid as the black line with a shaded region indicating the typical ± 0.3 dex scatter (e.g. Tacchella et al. 2016). All three classes of AGNs; RQAGN, HERG, and LINELERG scatter in large part below the main sequence, with

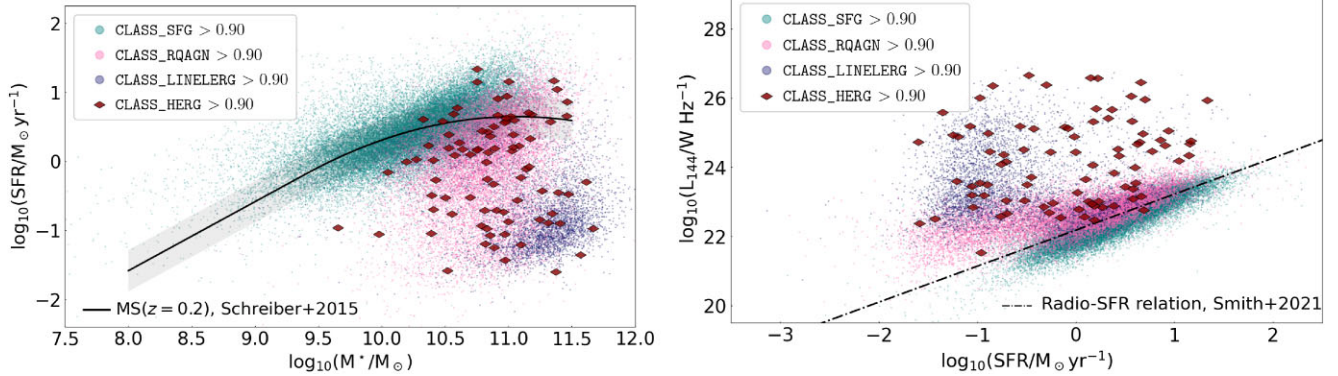


Figure 11. The distribution in the (Left:) SFR–stellar mass and (Right:) SFR– $L_{144\text{ MHz}}$ planes for the radio sources which have > 90 percent reliability in each class (SFG, RQAGN, HERG, and LINELERG). SFGs are indicated by the teal points, RQAGN as pink, and LINELERGs in dark blue. HERGs are overlaid as larger dark-red diamonds. In the left-hand panel, the ‘main sequence’ relation from Schreiber et al. (2015) evaluated at $z = 0.2$ along with a 0.3 dex scatter is indicated by the black line, for context. Similarly, in the right panel, the relationship between SFR and 144 MHz luminosity from Smith et al. (2021) is shown as the black dotted–dashed line in the right-hand panel. The SFR and stellar mass estimates come from the MPAJHU catalogue as an independent validation of our method.

some level of overlap. Our source distributions demonstrate that our probabilistic spectral source classifications broadly select objects with the expected range of stellar masses and SFRs for their physical class, measured through an independent catalogue. The distribution is also qualitatively similar to that seen in Gürkan et al. (2018) who used early LOFAR observations of the H-ATLAS NGP to classify $\approx 15\,000$ radio sources in a search for SFGs. By using Best & Heckman (2012) spectral classifications to isolate radio-AGN, and in conjunction with a BPT analysis applied to MPA-JHU emission line measurements (Brinchmann et al. 2004), the authors verified their SED-fitting–derived SFR measurements by showing that the classified sources fall as expected relative to the main sequence of SFGs (which in turn is well-aligned with the Elbaz et al. 2007 main-sequence relation). In the right-hand panel of Fig. 11, we show the $\log_{10}\text{SFR}$ versus $\log_{10}(L_{144})$ plane, combining the LOFAR $\log_{10}(L_{144})$ measurement (Hardcastle et al. 2023) with the MPA-JHU SFR, and colour-coding by our independently-measured source classifications. Our securely classified SFGs (90 per cent threshold) form a sequence in L_{144} – $\log_{10}\text{SFR}$ – the well-known radio–SFR relation (e.g. Smith et al. 2021, which has been overlaid as the black dotted–dashed line), whereas securely classified HERGs and LERGs scatter above the relation as expected, due to the excess in their radio luminosities as a consequence of their AGN activity. It is worth reiterating once again, that, although the HERG and LERG classes are indeed selected on the basis of a radio excess relative to the L_{144} predicted by their SFR (in our case their Balmer-corrected $H\alpha$ luminosity which directly correlates with SFR) the values of SFR in Fig. 11 have been derived independently in the MPA-JHU catalogue (Brinchmann et al. 2004), and thus serve as external validation of our analysis. Of particular interest in this panel are the RQAGN. These objects largely overlap with the main body of SFGs, but exhibit a tail of objects towards low stellar masses ($\log_{10}\text{SFR} \lesssim -1$) with radio luminosities elevated slightly above the radio–SFR relation. The nature of the radio emission in these objects is likely to be studied with increasing levels of interest as radio surveys continue to progress; for instance as LOFAR2.0 comes online (and in a number of pioneering works harnessing the international long baselines of LOFAR already; e.g. Morabito et al. 2022; Sweijen et al. 2023) imaging at 144 MHz will resolve $\sim \text{kpc}$ scales, potentially revealing the presence of either small-scale radio jets, or distributed, highly obscured star formation.

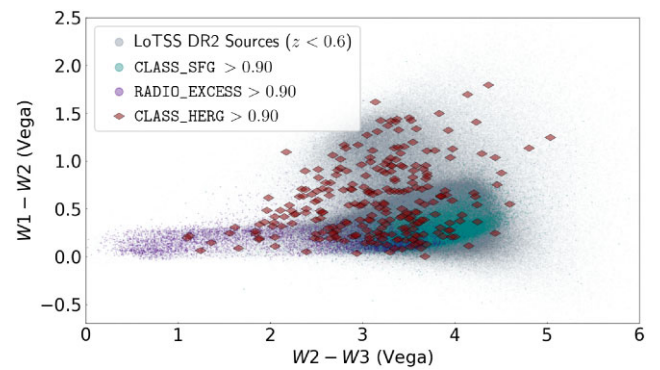


Figure 12. *WISE* colour–colour plot in Vega magnitudes for objects with good *WISE* photometry. The logarithmic density plot in grey shows 1.3 million DR2 objects with good photometry from Hardcastle et al. (2023). Colours indicate high-confidence SFGs, radio-excess sources, and the HERG subset of radio-excess objects.

4.5 Additional tests

4.5.1 *WISE* colour and luminosity validation

The *WISE* colour–colour plot is a widely used diagnostic for separating different classes of galaxies (e.g. Mateos et al. 2012; Gürkan, Hardcastle & Jarvis 2014). In Fig. 12, we show the colour–colour plot for galaxies in the DR2 optically identified sample (Hardcastle et al. 2023) with galaxies classified at high-confidence levels from this paper overlaid. It can be seen that star-forming galaxies and radio-excess objects both generally lie on the normal galaxy continuum with $0 < W1 - W2 < 0.5$ (i.e. near-infrared colours dominated by an old stellar population) but are clearly distinguished by their $W2 - W3$ colours, with SFGs having much redder $W2 - W3$ due to the importance of warm dust in their overall SEDs. HERGs clearly lie away from the bulk of the (presumably LERG) radio-excess population, which is expected due to the contribution of the AGN torus to $W3$, and was seen in the similar analysis of bright radio sources by Gürkan et al. (2014), although there the proportion of HERGs was far larger. Also as expected, some of our HERGs lie in the location generally populated by quasars and likely could be classified as quasars, Seyfert 1s or broad-line radio galaxies.

Importantly, many of the HERGs lie in the region generally populated by SFGs, showing that it is not safe to separate SFGs and radio-excess AGN by their *WISE* colours alone.

As shown by Hardcastle et al. (2023), we can also relate the *WISE* luminosity (absolute magnitude, here computed using a simple power-law extrapolation between *WISE* bands) to the radio luminosity for these sources, and this is shown in Fig. 13. Another validation of our methods is that objects that we classify as SFGs or radio-excess objects are very differently positioned on these plots, and behave in different ways for *W2* and *W3* absolute magnitudes. *W2* acts as a proxy of galaxy mass, and we see that radio-excess objects start to appear only in massive galaxies: a *W2*– L_{144} correlation appears for SFG because of the relation between galaxy mass and star formation rate. However, there is a clear region of overlap between SFG and radio-excess objects. By contrast, the two populations are very well separated on the *W3* plot where the correlation arises because of the relationship between dust luminosity and star-formation rate that gives rise to the well-known (far)-infrared/radio correlation (e.g. van der Kruit 1971; Yun, Reddy & Condon 2001; Jarvis et al. 2010; Smith et al. 2014). Radio-excess objects lie significantly above the main sequence of star formation as indicated by this correlation. Our emission-line classifications would allow us to calibrate a method of selecting radio-excess objects using only their *WISE* and radio properties, in a similar way to the *Herschel*-based approach of Hardcastle et al. (2016). On these plots HERGs tend to be found at high radio luminosities, but also at high *W3* absolute magnitudes, as noted by Gürkan et al. (2014).

4.5.2 Building on this framework: Specific [O III] luminosity of HERGs & LERGs validation

In the picture in which the HERG/LERG dichotomy is a transition between radiatively efficient and inefficient AGN activity occurring at a level of a few per cent of the Eddington rate (Best & Heckman 2012; Mingo et al. 2014; Hardcastle 2018), we expect proxies of L/L_{Edd} to show a transition in samples that have reliable optical emission-line classifications. Locating this transition could also help us to classify objects that do not exhibit all the lines needed for a full BPT classification using our method.

In radio-excess objects, we can take the luminosity in the [O III] line as a proxy for the radiative activity of the AGNs; standard linear corrections exist to turn this into an approximate bolometric luminosity (Heckman et al. 2004) although these of course only apply to radiatively efficient objects. We do not have the black hole mass estimates needed to estimate the Eddington luminosity, but can make use of the galaxy mass/black hole mass relation (e.g. Reines & Volonteri 2015) as a proxy of M_{BH} . In this analysis, we use the galaxy mass estimates from Hardcastle et al. (2023), which are derived from the DESI Legacy Survey and *WISE* photometry and so are superior to those based on SDSS photometry, but are not available for all sources.

Fig. 14 shows a histogram of the specific [O III] luminosity for radio-excess objects ($p > 0.95$) broken down into HERG and LINELERG classes (where we use $p = 0.8$ as the threshold for the emission-line classification). As expected (though not universally seen: Whittam et al. 2022), the objects we classify as HERGs have systematically high-specific [O III] luminosity, although there is substantial overlap, as seen by Best & Heckman (2012) and Kondapally et al. (in preparation), which is not surprising as many of the LERG are not significantly detected in this line. Moreover, the distribution for HERGs is at least centred in roughly the right

location. Take the critical accretion rate for radiatively efficient accretion to be $0.03\dot{M}_{\text{Edd}}$, and we know that $\dot{M}_{\text{Edd}} = 2.3 \times 10^{-9} M_{\text{BH}} \text{ yr}^{-1}$. The black hole mass to galaxy mass ratio is around 0.003 for elliptical galaxies (Kormendy & Ho 2013; Reines & Volonteri 2015), thus the critical accretion rate per galaxy mass is $\sim 2 \times 10^{-13} \text{ yr}^{-1}$, or $6.5 \times 10^{-21} \text{ s}^{-1}$. For an AGN efficiency of $\eta = 0.1$ and a bolometric correction of $B = 3500$ from Heckman et al. (2004), we expect the HERG/LERG transition at $6.5 \times 10^{-21} \eta M_{\odot} c^2 / B = 3.3 \times 10^{22} \text{ W } M_{\odot}^{-1}$, which is in line with what is observed. The HERG distribution should in theory only be 1.5–2 decades wide, but is of course broadened by the decade or so of intrinsic scatter in the black hole mass/galaxy mass relation, as well as uncertainties on efficiency factors, bolometric corrections, and the measurements of galaxy mass and luminosity themselves.

Fitting a KDE estimator to the histogram allows us to see that a naive Bayesian classifier would state that any object with $\log_{10}(L/M_{\odot}) > 23.5$ is most likely to be a HERG, and applying this to the unclassified sources in our data set, which typically do not have all the lines detected that are needed for a BPT classification, would allow us to select an additional ~ 80 objects that could be classified as HERGs. This modification to the scheme would effectively reinstate the ‘equivalent width of [O III]’ criterion for HERG selection used in some previous work (Laing et al. 1994), but now with an explicit physical motivation. Given the small numbers involved, we do not attempt to apply this classification to the present catalogue. The nature of the HERG and LERG in these samples, making use of this method, and the relationship between emission-line class and radio properties will be investigated in a future paper (Chilufya et al. in preparation).

5 SPECTROSCOPIC SOURCE DEMOGRAPHICS IN LOTSS DR2

5.1 Overall source demographics

We describe in this section, the demographic breakdown of the LoTSS DR2 sources which have been studied in this work. As we discuss in Section 3.4, one can construct a particular physical class of objects using different reliability criteria according to the need for purity or completeness of a given sample. This is especially straightforward for the physical-class reliabilities already provided in our catalogue, but we can also utilize our probabilistic RADIO_EXCESS measurements, and re-construct a class of radiative AGN (i.e. containing objects in either CLASS_RQAGN or CLASS_HERG) to represent numbers along the two independent axes with which we identify AGNs. As a demonstration of two potential approaches to purity and completeness, we conduct the following demographic analysis applying two different thresholds – a very high reliability threshold (> 99 per cent reliability) and a very permissive threshold (> 50 per cent reliability) requiring that objects must be classified into the same class for at least 99 per cent/50 per cent of the catalogue realizations.

Initially, we examine the breakdown of the subsets of objects likely to be radiative- and radio-excess AGN as a function of radio flux density and radio luminosity. These subsets are constructed by selecting objects with $\text{BPT_SFG} < 0.01$ (0.50) to capture all subclasses of radiative AGN, and by applying a reliability cut of 0.99 (0.50) in RADIO_EXCESS, respectively. These breakdowns are shown in Fig. 15. In the top left-hand panel, we show our most reliable measurements of the two diagnostics as a function of flux density (requiring that > 99 per cent of realizations satisfy each criterion). Below flux levels of ≈ 25 mJy the dominant population

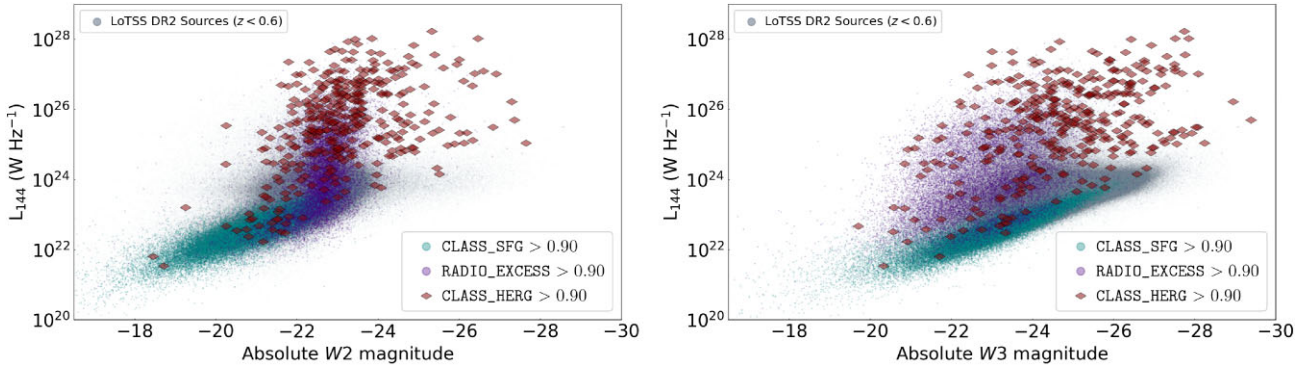


Figure 13. Radio luminosity as a function of (left) W2 and (right) W3 absolute magnitude. The logarithmic density plot in grey shows 1.0 million DR2 sources with $z < 0.6$ as a comparison population. Colours as in Fig. 12.

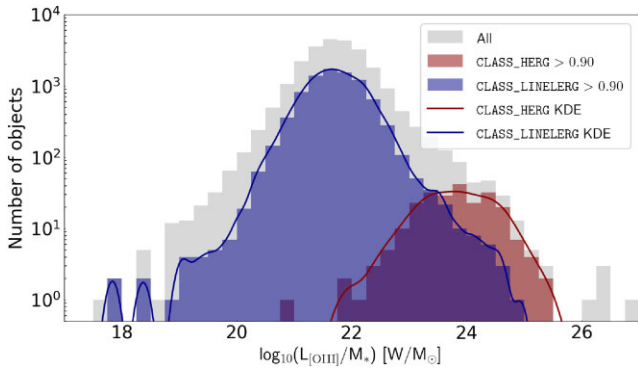


Figure 14. The distribution of the specific [O III] luminosity for 19 331 radio-excess objects with mass estimates and [O III] measurements in our sample, colour coded by their emission-line classification status, if any. Histograms show the raw source counts, solid lines show kernel density estimates smoothing over the observed distribution.

consists of objects which do not satisfy either of these criteria 99 per cent of the time, whereas at the highest fluxes > 25 mJy, radio-excess sources dominate the source counts. Radiative AGNs make up roughly 20 per cent of the population at all flux levels. In the top right-hand panel, we show the equivalent plot this time applying a liberal 50 per cent threshold. In this case we see radio excess objects constituting approximately 30 per cent of our sample up to a flux level of ≈ 3 mJy, before rising smoothly to make up almost 100 per cent of the sample above flux densities ≈ 40 mJy.

In the lower-left panel of Fig. 15, as a function of log luminosity, we see similar behaviour in that below $\log_{10}(L_{144}/W \text{ Hz}^{-1}) \approx 24.5$, the majority of sources are neither radiative AGN or Radio-Excess AGN in 99 per cent of realizations. Again the radiative AGN compose an approximately constant 20 per cent of sources below $\log_{10}(L_{144}/144 \text{ W Hz}^{-1}) \approx 23$, reaching approximately 100 per cent by $\log_{10}(L_{144}/W \text{ Hz}^{-1}) \approx 27.5$. This reflects the behaviour we expect for radio-selected samples, as we discuss in Section 4.3. In the lower right panel, we apply our liberal 50 per cent threshold, finding that only below $\log_{10}(L_{144}/W \text{ Hz}^{-1}) \approx 23$ do sources which are neither radiative AGNs or radio-excess AGNs dominate the sample. At approximately the same luminosity radiative AGNs begin to dominate the fractional counts, rising steeply to ≈ 100 per cent by $\log_{10}(L_{144}/W \text{ Hz}^{-1}) \approx 24.5$. Radio excess object also reach ≈ 100 per cent but only at the higher luminosity of $\log_{10}(L_{144}/W \text{ Hz}^{-1}) \approx 27.5$.

We next consider four non-overlapping physical classes of sources, classified as described mainly in Section 3. In addition to CLASS_SFG, CLASS_RQAGN, and CLASS_HERG, we combine our recorded diagnostics to define the final physical class of objects: LERGs (low-excitation radio galaxies). To reconstruct the full LERG class requires that we also consider the information gained from the *lack* of measured emission lines. As such, we combine the emission-line LERG class (CLASS_LINEHERG) with those objects with a RADIO_EXCESS reliability greater than our chosen threshold, and that show a CLASS_HERG reliability less than 1 minus this threshold (i.e. we want to select the sources least likely to lie in CLASS_HERG, as by definition LERGs are radio-excess sources which are not highly ionized).

In Fig. 16, we show three rows of panels which record the demographic break down of securely classified (> 99 per cent reliability; left-hand column) and ‘liberally’ classified (> 50 per cent reliability; right-hand column) sources in each of the four physical classes described above. Objects that do not meet the threshold in any one class are recorded as unclassified (‘UNC’).

In the upper panels, we show the demographics as a function of the radio flux at 144 MHz; $\log_{10}(F_{144})$. In the left-hand panel, as a result of our strict criterion to ensure high purity of physical classes, the majority of sources appear in the UNC subset at all flux densities. Amongst securely classified subsets the LERGs dominate above ≈ 10 mJy, reaching ≈ 50 per cent of sources at the highest flux densities (≈ 315 mJy). Below ≈ 10 mJy SFGs are the dominant population, peaking at ≈ 2.2 mJy. RQAGN make up approximately 10 per cent of the population below ≈ 10 mJy (the same threshold at which LERGs become more numerous than SFGs) and gradually decline above this, reaching roughly zero by ≈ 40 mJy. The HERG class are barely visible on this plot, reaching only a few per cent of the fractional counts by flux densities above ≈ 100 mJy. In the right-hand panel where we apply a liberal threshold of 50 per cent very similar behaviour is seen across all physical classes, only with fractional counts scaled up by about 10 per cent. This results in the unclassified class ‘UNC’ dropping substantially to lie below the individual physical classes down to around ≈ 1 mJy, more in-line with previously reported demographics, where the doubt on any individual classification has not been quantified.

In the central panels of Fig. 16, we present the demographic breakdown of sources as a function of luminosity, plotted as $\log_{10}(L_{144}/W \text{ Hz}^{-1})$. In the left-hand panel showing our most securely classified sources, it is interesting to note that across the entirety of the central luminosity range; $22.0 < \log_{10}(L_{144}/W \text{ Hz}^{-1}) < 26.5$, SFGs dominate the source counts. On the right of the

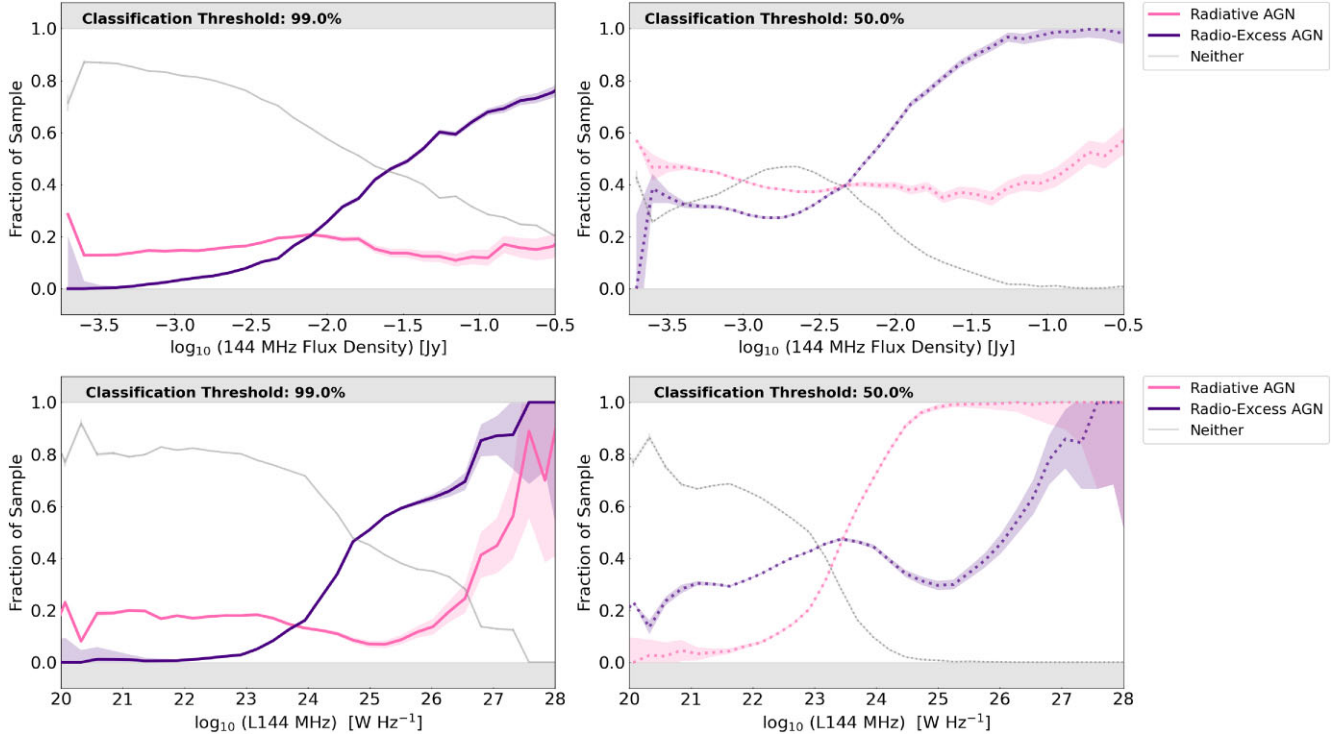


Figure 15. The fraction of sources identified as having a > 99 (> 50) per cent reliability of hosting a radiative AGN, having a radio excess, or neither, are shown as pink, purple, and grey, respectively (as indicated in the legend) as a function of 144-MHz flux density (top) and luminosity (bottom panel). Error bars have been calculated assuming Poisson statistics.

plot, at the highest luminosities, LERGs and HERGs dominate the source counts, exceeding unclassified objects at $\log_{10}(L_{144}/\text{W Hz}^{-1}) \approx 26.25$ and $\log_{10}(L_{144}/\text{W Hz}^{-1}) \approx 27.0$, respectively. This panel in particular goes to show the importance of incorporating a ‘quantified doubt’ when reporting demographics, or examining the behaviour of a particular class of objects – particularly across the luminosity range where it happens that most LoTSS-Wide detections lie. In the right-hand panel, we again see behaviour of the physical classes very similar to the most secure classifications – for SFGs and RQAGN the fractional counts again seem to simply scale up by about 10 per cent, meanwhile the LERGs and HERGs are slightly more complex. Unclassified objects drop dramatically to being the dominant class only for ≈ 2 decades, and are only very slightly (< 10 per cent) more numerous than the LERGs in this range. The LERG class interestingly receives a boost across all luminosities up to about $\log_{10}(L_{144}/\text{W Hz}^{-1}) \approx 26$ whereby their counts actually drop more so than for the securely classified subset, and appear to shift into the HERG class instead. This is a potentially interesting observation if this ambiguity between the HERG and LERG classes is uncaptured in other works.

Finally, in the lower panels of Fig. 16, we show our demographic breakdown as a function of spectroscopic redshift, z (as recorded in the Portsmouth catalogue). While it is interesting to examine these numbers, our combination of spectroscopic catalogues that target different classes of objects and across two different redshift ranges (as well the fact that our sample comprises only a few per cent of LoTSS sources) means the demographics as a function of redshift are unlikely to be representative of the full population of galaxies in the Universe.

For our most securely classified sources (in the left-hand panel) only at the very lowest redshifts does any one physical class (the

SFGs) dominate above unclassified sources. At $z > 0.05$, unclassified sources are by far the most numerous class, reaching 75 per cent of the sample by $z \approx 0.15$, rising almost to 100 per cent by the highest redshifts in this work ($z \approx 0.55$). Considering only the securely classified subsets, SFGs drop rapidly, reaching zero by $z \approx 0.255$. RQAGN constitute ≈ 17 per cent of the population out to $z \approx 0.2$, with a secondary bump at $z = 0.4\text{--}0.5$. LERGs are at very low levels across most redshifts, but reach ≈ 17 per cent across the $z = 0.25\text{--}0.4$ range. HERGs are likewise at very low (< 10 per cent) levels across the entire redshift range. In the right-hand panel which shows our more liberally selected subsets, we find unclassified sources drop to a level comparable with LERGs across the majority of redshifts ($z \gtrsim 0.2$). The LERGs are once again the most dramatically boosted class by lowering the reliability threshold, meanwhile SFGs and RQAGN scale upwards by ≈ 10 per cent, and while this may also be true of the HERGs, they remain undetectable in the fractional counts until a rise at $z \gtrsim 0.35$ to a few per cent of the total counts.

6 CONCLUSIONS

We have produced probabilistic classifications for the ~ 4 per cent of LoTSS DR2 (Shimwell et al. 2022) sources for which spectroscopic measurements exist in the Portsmouth catalogue (Thomas et al. 2013), based on their analysis of SDSS and BOSS spectra. We have done this by combining the spectroscopic and low-frequency radio information with extensive Monte Carlo simulations, in which we have produced multiple realizations of the input catalogues in a manner that accounts for the measured uncertainties. Since our classifications are probabilistic, they can be used to produce samples of sources tailored according to the scientific requirements of the end user, whether that be high purity (e.g. including only

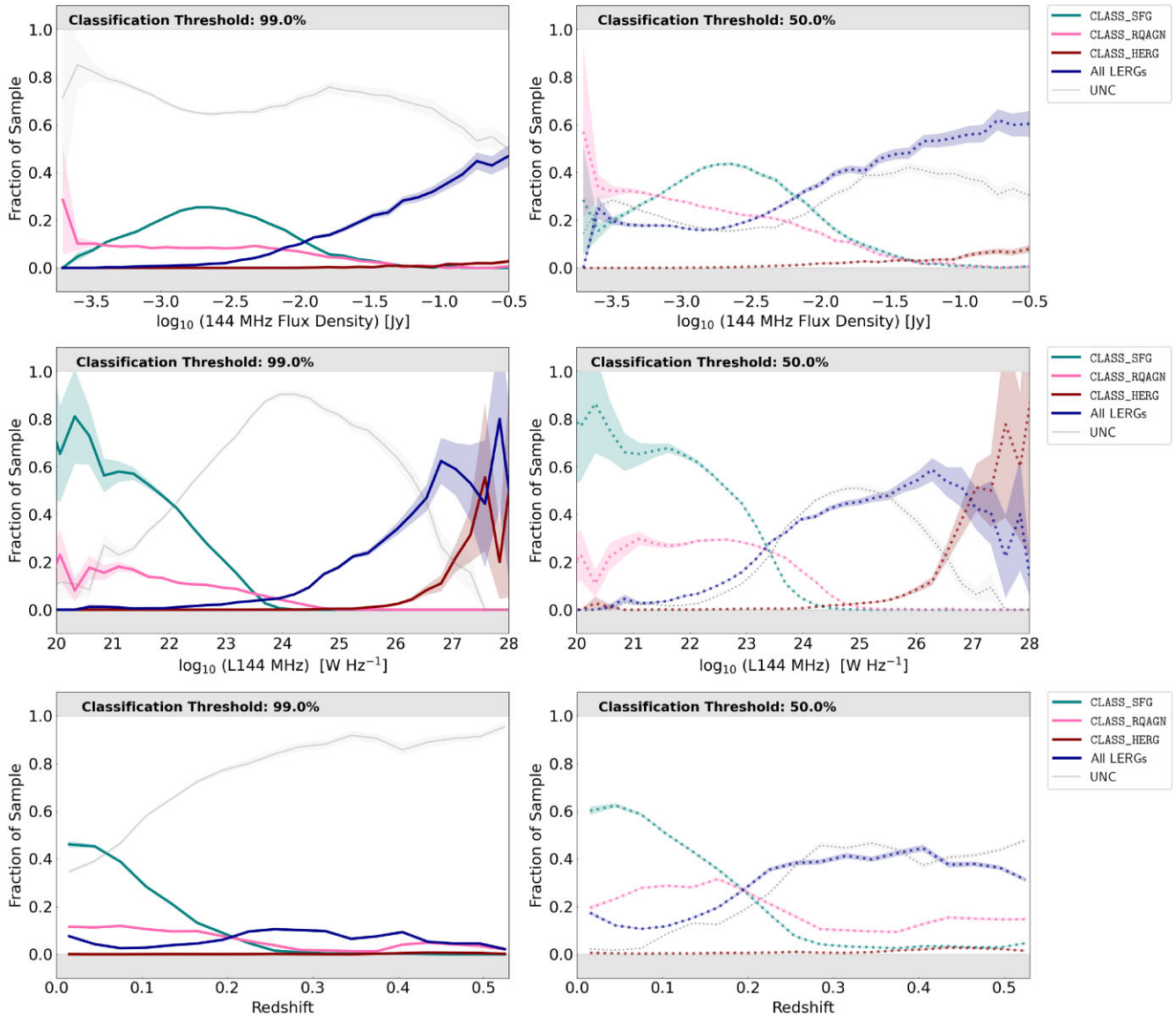


Figure 16. The fraction of sources with > 99 (> 50) per cent reliability classifications as SFG, RQAGN, HERGs, LINELERGs are shown as teal, pink, red, and blue symbols, respectively (as indicated in the legend) as a function of 144 MHz flux density (top), 144 MHz luminosity (middle), and redshift (bottom). In each panel, sources that do not reach the classification threshold in any of the four categories are shown as unclassified ('UNC', and indicated in grey). Error bars have been estimated using Poisson statistics.

the most reliably classified High-Excitation Radio Galaxies) or completeness (e.g. including any source that may have a radio excess).

Using a 90 per cent reliability threshold, in the LoTSS DR2 sample we have identified 38 588 radio-excess AGNs (i.e. those which have an excess of 144 MHz emission over that which would be expected on the basis of their Balmer-corrected $H\alpha$ line alone), and identified 12 648 LINELERGs and 362 HERGs among them. Similarly, we have been able to identify 38 729 SFGs and 18 726 RQAGN. Our catalogue contains the largest sample of spectroscopically classified radio sources to date, far surpassing previous works in the literature (e.g. Best & Heckman 2012; Sabater et al. 2019; Whittam et al. 2022).

We have validated our method by comparing our results with the wider literature, including comparisons with previous catalogues such as Best & Heckman (2012), with other diagnostics such as the

Excitation Index (Buttiglione et al. 2010), and with our expectations based on our previous investigations of the radio source population (e.g. Best et al. 2023; Das et al. 2024).

Although our method has produced classifications for only around 4 per cent of the LoTSS DR2 sources, our method has been designed to be directly applicable to the observations of the WEAVE-LOFAR survey. WEAVE-LOFAR (Smith et al. 2016) will – in 2024 – begin to produce more than a million spectra of sources identified purely on the basis of their 144 MHz emission, and include spectroscopy of every radio source detected in the LoTSS deep fields (Duncan et al. 2021; Kondapally et al. 2021; Sabater et al. 2021; Tasse et al. 2021).

ACKNOWLEDGEMENTS

The authors would like to thank the referee for a constructive report that improved the quality of this paper. ABD, DJBS, MJH, and MIA

acknowledge support from the UK Science and Technology Facilities Council (STFC) grant ST/V000624/1. DJBS also acknowledges STFC support from grant ST/Y001028/1. PNB is grateful for support from the UK STFC via grant ST/V000594/1. We're also encouraged by our national open access policies to add a comment to the acknowledgements on the arXiv version: 'For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission. RK is grateful for support from the UK STFC via grant ST/V000594/1. MIA acknowledges support from the UK Science and Technology Facilities Council (STFC) studentship under the grant ST/V506709/1. SD acknowledges support from a Science and Technology Facilities Council (STFC) studentship via grant ST/W507490/1. SS acknowledges support from the UK Science and Technology Facilities Council (STFC) via grant ST/X508408/1. KJD acknowledges support from the STFC through an Ernest Rutherford Fellowship (grant number ST/W003120/1).

LOFAR is the Low Frequency Array, designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the ILT foundation under a joint scientific policy. The ILT resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Université d'Orléans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK; Ministry of Science and Higher Education, Poland; The Istituto Nazionale di Astrofisica (INAF), Italy.

This research made use of the Dutch national e-infrastructure with support of the SURF Cooperative (e-infra 180169) and the LOFAR e-infra group. The Jülich LOFAR Long Term Archive and the German LOFAR network are both coordinated and operated by the Jülich Supercomputing Centre (JSC), and computing resources on the supercomputer JUWELS at JSC were provided by the Gauss Centre for Supercomputing e.V. (grant CHTB00) through the John von Neumann Institute for Computing (NIC).

This research made use of the University of Hertfordshire high-performance computing facility and the LOFAR-UK computing facility located at the University of Hertfordshire (<https://uhhpc.hert.ac.uk>) and supported by STFC [ST/P000096/1], and of the Italian LOFAR IT computing infrastructure supported and operated by INAF, and by the Physics Department of Turin University (under an agreement with Consorzio Interuniversitario per la Fisica Spaziale) at the C3S Supercomputing Centre, Italy.

DATA AVAILABILITY

This paper is based on LoTSS DR2 optical identifications. The optical identification catalogue, together with the classifications produced in this work (and the documentation necessary to use them) are available to download from https://lofar-surveys.org/dr2_release.html. Other data presented in the paper are available upon reasonable request to the corresponding author.

REFERENCES

- Aihara H. et al., 2011, *ApJS*, 193, 29
 Algera H. S. B. et al., 2023, *MNRAS*, 518, 6142
 Bañados E. et al., 2018, *Nature*, 553, 473
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
 Barrufet L. et al., 2023, *MNRAS*, 522, 3926
 Becker R. H., White R. L., Helfand D. J., 1995, *ApJ*, 450, 559
 Bell E. F., 2003, *ApJ*, 586, 794
 Best P. N., Heckman T. M., 2012, *MNRAS*, 421, 1569 (BH12)
 Best P. N., Kauffmann G., Heckman T. M., Ivezić Ž., 2005, *MNRAS*, 362, 9
 Best P. N., Ker L. M., Simpson C., Rigby E. E., Sabater J., 2014, *MNRAS*, 445, 955
 Best P. N. et al., 2023, *MNRAS*, 523, 1729
 Bouwens R. J. et al., 2016, *ApJ*, 833, 72
 Bouwens R. J., Illingworth G., Ellis R. S., Oesch P., Stefanon M., 2022, *ApJ*, 940, 55
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151
 Buttiglione S., Capetti A., Celotti A., Axon D. J., Chiaberge M., Macchetto F. D., Sparks W. B., 2010, *A&A*, 509, A6
 Cid Fernandes R., Stasińska G., Schlickmann M. S., Mateus A., Vale Asari N., Schoenell W., Sodré L., 2010, *MNRAS*, 403, 1036
 Cid Fernandes R., Stasińska G., Mateus A., Vale Asari N., 2011, *MNRAS*, 413, 1687
 Condon J. J., 1992, *ARA&A*, 30, 575
 Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, *AJ*, 115, 1693
 Dalton G. et al., 2012, in McLean I. S., Ramsay S. K., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P
 Das S. et al., 2024, *MNRAS*, 531, 977
 Duncan K. J. et al., 2021, *A&A*, 648, A4
 Dunlop J. S. et al., 2017, *MNRAS*, 466, 861
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72
 Elbaz D. et al., 2007, *A&A*, 468, 33
 Farina E. P. et al., 2022, *ApJ*, 941, 106
 Fioc M., Rocca-Volmerange B., 1997, *A&A*, 326, 950
 Fudamoto Y. et al., 2021, *Nature*, 597, 489
 Gruppioni C. et al., 2013, *MNRAS*, 432, 23
 Gürkan G., Hardcastle M. J., Jarvis M. J., 2014, *MNRAS*, 438, 1149
 Gürkan G. et al., 2015, *MNRAS*, 452, 3776
 Gürkan G. et al., 2018, *MNRAS*, 475, 3010
 Hardcastle M., 2018, *Nat. Astron.*, 2, 273
 Hardcastle M. J., Croston J. H., 2020, *New Astron. Rev.*, 88, 101539
 Hardcastle M. J., Evans D. A., Croston J. H., 2007, *MNRAS*, 376, 1849
 Hardcastle M. J. et al., 2016, *MNRAS*, 462, 1910
 Hardcastle M. J. et al., 2023, *A&A*, 678, A151
 Hatcher C. et al., 2021, *AJ*, 162, 65
 Heckman T. M., 1980, *A&A*, 87, 152
 Heckman T. M., Best P. N., 2014, *ARA&A*, 52, 589
 Heckman T. M., Kauffmann G., Brinchmann J., Charlot S., Tremonti C., White S. D. M., 2004, *ApJ*, 613, 109
 Hickox R. C., Alexander D. M., 2018, *ARA&A*, 56, 625
 Hill G. J. et al., 2008, in Kodama T., Yamada T., Aoki K. eds, ASP Conf. Ser. Vol. 399, Panoramic Views of Galaxy Formation and Evolution. Astron. Soc. Pac., San Francisco, p. 115
 Hine R. G., Longair M. S., 1979, *MNRAS*, 188, 111
 Hopkins A. M., Beacom J. F., 2006, *ApJ*, 651, 142
 Jarvis M. J. et al., 2010, *MNRAS*, 409, 92
 Kauffmann G. et al., 2003, *MNRAS*, 341, 33
 Kauffmann G., Heckman T. M., Best P. N., 2008, *MNRAS*, 384, 953
 Kennicutt R. C., Evans N. J., 2012, *ARA&A*, 50, 531
 Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, 556, 121
 Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, *MNRAS*, 372, 961
 Kistler M. D., Yüksel H., Beacom J. F., Hopkins A. M., Wyithe J. S. B., 2009, *ApJ*, 705, L104
 Kondapally R. et al., 2021, *A&A*, 648, A3
 Kondapally R. et al., 2022, *MNRAS*, 513, 3742
 Kormendy J., Ho L. C., 2013, *ARA&A*, 51, 511
 Kroupa P., 2001, *MNRAS*, 322, 231
 Laing R. A., Riley J. M., Longair M. S., 1983, *MNRAS*, 204, 151

- Laing R. A., Jenkins C. R., Wall J. V., Unger S. W., 1994, in Bicknell G. V., Dopita M. A., Quinn P. J. eds, ASP Conf. Ser. Vol. 54, The Physics of Active Galaxies. Astron. Soc. Pac., San Francisco, p. 201
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Magliocchetti M. et al., 2002, *MNRAS*, 333, 100
- Magnelli B., Elbaz D., Chary R. R., Dickinson M., Le Borgne D., Frayer D. T., Willmer C. N. A., 2011, *A&A*, 528, A35
- Magnelli B. et al., 2013, *A&A*, 553, A132
- Martínez-Sansigre A., Rawlings S., Lacy M., Fadda D., Marleau F. R., Simpson C., Willott C. J., Jarvis M. J., 2005, *Nature*, 436, 666
- Mateos S. et al., 2012, *MNRAS*, 426, 3271
- Mazzucchelli C. et al., 2017, *ApJ*, 849, 91
- Mingo B., Hardcastle M. J., Croston J. H., Dicken D., Evans D. A., Morganti R., Tadhunter C., 2014, *MNRAS*, 440, 269
- Morabito L. K. et al., 2022, *MNRAS*, 515, 5758
- Narayan R., Yi I., 1995, *ApJ*, 452, 710
- Perna M. et al., 2015, *A&A*, 583, A72
- Reines A. E., Volonteri M., 2015, *ApJ*, 813, 82
- Robertson B. E., Ellis R. S., 2012, *ApJ*, 744, 95
- Sabater J. et al., 2019, *A&A*, 622, A17
- Sabater J. et al., 2021, *A&A*, 648, A2
- Sarzi M. et al., 2017, Astrophysics Source Code Library, record ascl:1708.012
- Schreiber C. et al., 2015, *A&A*, 575, A74
- Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
- Shimwell T. W. et al., 2017, *A&A*, 598, A104
- Shimwell T. W. et al., 2019, *A&A*, 622, A1
- Shimwell T. W. et al., 2022, *A&A*, 659, A1
- Smith D. J. B. et al., 2014, *MNRAS*, 445, 2232
- Smith D. J. B. et al., 2016, in Reylé C., Richard J., Cambrésy L., Deleuil M., Pécontal E., Tresse L., Vauglin I. eds, SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. p. 271, preprint ([arXiv:1611.02706](https://arxiv.org/abs/1611.02706))
- Smith D. J. B. et al., 2021, *A&A*, 648, A6
- Smolčić V. et al., 2017, *A&A*, 602, A1
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253
- Sweijen F. et al., 2023, *A&A*, 671, A85
- Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack Joel R., 2016, *MNRAS*, 457, 2790
- Tadhunter C. N., Morganti R., Robinson A., Dickson R., Villar-Martin M., Fosbury R. A. E., 1998, *MNRAS*, 298, 1035
- Tasse C. et al., 2021, *A&A*, 648, A1
- Thomas D. et al., 2013, *MNRAS*, 431, 1383
- Tremonti C. A. et al., 2004, *ApJ*, 613, 898
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- van der Kruit P. C., 1971, *A&A*, 15, 110
- Wang F. et al., 2021, *ApJ*, 907, L1
- Weedman D. W., 1977, *ARA&A*, 15, 69
- Whitaker K. E., Pope A., Cybulski R., Casey C. M., Popping G., Yun M. S., 2017, *ApJ*, 850, 208
- Whittam I. H. et al., 2022, *MNRAS*, 516, 245
- Williams W. L. et al., 2019, *A&A*, 622, A2
- Yang J. et al., 2020, *ApJ*, 897, L14
- York D. G. et al., 2000, *AJ*, 120, 1579
- Yun M. S., Reddy N. A., Condon J. J., 2001, *ApJ*, 554, 803
- Zavala J. A. et al., 2021, *ApJ*, 909, 165

APPENDIX A: EXAMPLE TABLE AND SUBSETS

In Table A1, we present the first few rows of our table to demonstrate how objects will appear. Objects 1 and 2 are fully classified sources, meanwhile objects 3 and 4 demonstrate the information available for sources that do not exhibit 4 BPT emission lines, but for which we have measured a radio excess. Non-overlapping subsets of the BPT diagram are as follows: BPT_SFG + BPT_SEY + BPT_LIN must = 1

To reconstruct the subset of objects above the Kewley et al. (2006) maximal starburst line: $AGN_{Kewley} = BPT_SEY + BPT_LIN - BPT_COMP$

For clarity, the subsets BPT_CSEY and BPT_CLIN are subsets of BPT_COMP. $BPT_CSEY + BPT_CLIN = BPT_COMP$

As discussed in the text, to reconstruct a LERG class which incorporates all objects with a radio excess that have not been classed as HERGs, we apply a threshold of choice e.g. 90 per cent and compute the following: $RADIO_EXCESS > 0.90$ and $CLASS_HERG < 0.10$

Table A1. An example of the machine-readable data table we provide with this work. Columns are described in Table 2. Four objects are included, two for which BPT analysis and classifications have been derived (objects 1 and 2), and two for which radio-excess analysis was carried out, but BPT lines were not available for further analysis (and hence full classification; objects 3 and 4).

Column	Object 1	Object 2	Object 3	Object 4
Source_Name	ILTJ000001.80+193250.3	ILTJ000002.57+200308.5	ILTJ000004.53+220436.5	ILTJ000007.23+332358.6
SPECOBJID	6946961409339748352	6898653816788021248	6879497850414886912	8043642899621670912
TOT_F144 (mJy)	3.04097	3.03803	5.60489	3.21012
ID_RA	0.00738	0.01082	0.01880	0.03024
ID_DEC	19.54727	20.05232	22.07691	33.39955
z_best	0.09049	0.27949	0.47511	0.48936
REDSHIFT	0.09067	0.27973	0.47530	0.48958
CLASS_z_WARNING	0.000	0.000	0.000	0.000
RADIO_EXCESS	0.000	0.053	0.569	0.999
BALMER_CORR_WARNING	0.000	0.000	0.000	1.000
BPT_SFG	0.333	0.355	–	–
BPT_COMP	0.667	0.642	–	–
BPT_SEY	0.090	0.027	–	–
BPT_LIN	0.577	0.618	–	–
BPT_CSEY	0.090	0.027	–	–
BPT_CLIN	0.577	0.615	–	–
BPT_ML	Y	Y	–	–
zscore	0.55777	0.50295	–	–
CLASS_SFG	0.333	0.350	–	–
CLASS_RQAGN	0.667	0.597	–	–
CLASS_HERG	0.000	0.000	–	–
CLASS_LINELERG	0.000	0.053	–	–

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.