The Astrophysical Journal, 978:101 (16pp), 2025 January 01 https://doi.org/10.3847/1538-4357/ad957b

HETDEX-LOFAR Spectroscopic Redshift Catalog*

Maya H. Debski[1], Gregory R. Zeimann[2,3], Gary J. Hill[1,3], Donald P. Schneider[4,5], Leah Morabito[6,7], Gavin Dalton[8,9], Matt J. Jarvis[8], Erin Mentuch Cooper[1,3], Robin Ciardullo[4,5], Eric Gawiser[10], and Nika Jurlin[1]
[1] Department of Astronomy, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA
[2] Hobby Eberly Telescope, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX, 78712, USA
[3] McDonald Observatory, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA
[4] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[5] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[6] Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
[7] Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
[8] Oxford Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK
[9] RALSpace, STFC Rutherford Appleton Laboratory, Didcot OX11 0QX, UK
[10] Department of Physics and Astronomy, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA
Received 2024 June 26; revised 2024 November 8; accepted 2024 November 18; published 2024 December 27

Abstract

We combine the power of blind integral field spectroscopy from the Hobby–Eberly Telescope (HET) Dark Energy Experiment (HETDEX) with sources detected by the Low Frequency Array (LOFAR) to construct the HETDEX-LOFAR Spectroscopic Redshift Catalog. Starting from the first data release of the LOFAR Two-metre Sky Survey, including a value-added catalog with photometric redshifts, we extracted 28,705 HETDEX spectra. Using an automatic classifying algorithm, we assigned each object a star, galaxy, or quasar label along with a velocity/redshift, with supplemental classifications coming from the continuum and emission-line catalogs of the internal, fourth data release from HETDEX (HDR4). We measured 9087 new redshifts; in combination with the value-added catalog, our final spectroscopic redshift sample is 9710 sources. This new catalog contains the highest substantial fraction of LOFAR galaxies with spectroscopic redshift information; it improves archival spectroscopic redshifts and facilitates research to determine the [O II] emission properties of radio galaxies from $0.0 < z < 0.5$, and the Lyα emission characteristics of both radio galaxies and quasars from $1.9 < z < 3.5$. Additionally, by combining the unique properties of LOFAR and HETDEX, we are able to measure star formation rates (SFRs) and stellar masses. Using the Visible Integral-field Replicable Unit Spectrograph, we measure the emission lines of [O III], [Ne III], and [O II] and evaluate line-ratio diagnostics to determine whether the emission from these galaxies is dominated by active galactic nuclei or star formation and fit a new SFR–$L_{150\mathrm{MHz}}$ relationship.

Unified Astronomy Thesaurus concepts: Redshift surveys (1378); Galaxy evolution (594); Radio continuum emission (1340); Radio active galactic nuclei (2134)

Materials only available in the online version of record: machine-readable table

1. Introduction

For several decades, extragalactic radio surveys were a powerful probe of the distant Universe. In fact, until the mid-1990s, they served as an effective method for finding high-redshift galaxies through optical identification of ultrasteep radio sources, a poorly understood type of diffuse radio source characterized by power-law spectra (e.g., O. B. Slee et al. 2001; L. Feretti et al. 2012; A. Whyley et al. 2024). Recent radio surveys have reached sub-millijansky flux density levels, providing the framework to make these surveys a means of identifying star-forming galaxies. Previous studies have demonstrated a tight correlation between low-frequency radio continuum, which is dominated by synchrotron emission of relativistic electrons produced by supernovae, and the far-infrared (FIR) flux of galaxies (e.g., M. S. Yun et al. 2001). FIR emission is an established strong indicator of star formation rate (SFR; M. S. Yun et al. 2001), thereby legitimizing the idea that radio continuum can act as a tracer of SFR (e.g., J. J. Condon et al. 2002; M. Pannella et al. 2009; V. Heesen et al. 2014; G. R. Davies et al. 2017; D. J. B. Smith et al. 2021).

Radio surveys have played a small part in investigating star formation history, as they are generally limited by their sensitivity; however, star-forming galaxies become increasingly important at fainter flux densities and totally dominate the source counts below $\simeq$0.1 mJy at 1.4 GHz (e.g., P. Padovani et al. 2011, 2015; M. Bonzini et al. 2013), opening the door for new, more sensitive radio surveys to push to the forefront of SFR investigations (G. De Zotti et al. 2019). In particular, the introduction of the Square Kilometer Array (SKA; K. Grainge et al. 2017) with its square kilometer collecting surface and large range of frequencies (between 50 MHz and 24 GHz) will extend the flux density limit more than 3 orders of magnitude. The science of the SKA will be used to explore areas such as strong-field tests with pulsars and black holes, cosmic dawn and the epoch of reionization, cosmology and dark energy, the origin and evolution of cosmic magnetism, galaxy evolution probed by neutral hydrogen, the cradle of life and astrobiology, and galaxy and cluster evolution (G. De Zotti et al. 2019). To

* Based on observations obtained with the Hobby–Eberly Telescope, which is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München, and Georg-August-Universität Göttingen.

prepare for this new era of radio surveys, SKA will work alongside Pathfinders, like the International Low Frequency Array (LOFAR) Telescope (ILT), to contribute scientific and technical developments for direct use by SKA.

The combination of multiple massive surveys at different wavelengths enables scientific projects otherwise unachievable (e.g., D. J. B. Smith et al. 2016; W. L. Williams et al. 2019). The LOFAR Two-metre Sky Survey (LoTSS) is a sensitive, high-resolution (120–168 MHz, centered at 150 MHz) survey that has already collected millions of sources and is advancing our understanding of the formation and growth of massive black holes (e.g., P. N. Best et al. 2014; B. Mingo et al. 2019; J. Sabater et al. 2019; B. Mingo et al. 2022; M. Yue et al. 2023), the evolution of galaxy clusters (e.g., B. P. Venemans et al. 2007; D. Wylezalek et al. 2013; A. Botteon et al. 2020; R. Timmerman et al. 2022, 2024), and the properties of high-redshift radio sources (e.g., A. J. Gloudemans et al. 2021; C. M. Cordun et al. 2023). However, many of these scientific forefronts require optical counterparts for multiwavelength matching, as well as a robust set of distances or redshifts.

Hobby–Eberly Telescope (HET) Dark Energy Experiment (HETDEX; K. Gebhardt et al. 2021; G. J. Hill et al. 2021) is blind-spectroscopic survey conducted with the wide-field upgraded HET (L. W. Ramsey et al. 1998; G. J. Hill et al. 2021) using the Visible Integral Field Replicable Unit Spectrograph (VIRUS; G. J. Hill et al. 2021).[11] HETDEX aims to measure the expansion history of the Universe at $z \simeq 2.5$ by detecting and mapping the spatial distribution of about a million Lyα emitting galaxies (LAEs). The redshift range for LAE detection is $1.9 < z < 3.5$ over a total of $\sim$540 deg^2 (11 Gpc^3 comoving volume) including $\sim$400 deg^2 in the HETDEX Spring Field. The internal data release 4 includes 67.48 deg^2 of spectroscopic observations in the Spring Field with wavelengths spanning 3470–5540 Å. This amounts to 523 million fiber spectra. In the LOFAR/HETDEX overlap region, we aim to produce a new, large spectroscopic radio-source catalog that would facilitate breakthroughs in the study of galaxy protoclusters, emission-line properties of radio galaxies, and even radio-loud stars.

The combination of wide-field photometric surveys for spectral energy distribution information and optical spectroscopic follow-up from HETDEX/VIRUS offers a characterization of key physical parameters, the most important of which being distance as represented by spectroscopic redshift (e.g., Sloan Digital Sky Survey; A. Almeida et al. 2023). The combination of HETDEX and LOFAR will also allow the measurement of SFR and stellar mass because of the VIRUS sensitivity. Both the LOFAR radio selection and HETDEX optical spectroscopic identifications are sensitive to SFR and active galactic nuclei (AGN) activity. Using VIRUS, we can measure the emission lines of [O III], Hβ, [Ne III], and [O II]. Emission-line diagnostics like those of [O III]/Hβ, [Ne III]/[O II], R_{23}, and [O III]/[O II] can provide information on a system's metallicity and ionization parameter, and discriminate between excitation by AGN or star formation. Moreover, by comparing these line ratios as a function of stellar mass with those determined for galaxies found via selection methods, we can ultimately investigate how similar or dissimilar populations of galaxies are.

In Section 2, we describe the observations, data sets, and tools required to identify HETDEX counterparts to LOFAR sources and analyze the objects. Section 3 focuses on the redshift analysis of the sample, including a description of the classification code Diagnose that was developed for the HET VIRUS Parallel Survey (HETVIPS) and its application in this catalog. We also describe the matching process for the sources within both the HETDEX survey and LoTSS, as well as the criteria for determining whether the matches were accurate. This section also features a broad overview of the redshift results alongside the overall catalog breakdown and selected results from the catalog. In addition, we examine star formation in a subsample of galaxies in Section 4. Finally, we discuss possible scientific applications and uses for this data set in Section 5. Throughout this work, we use flat ΛCDM cosmological parameters $H_0 = 67.66$ km Mpc^{-1} s^{-1} and $\Omega_{m,0} = 0.30966$ (Planck Collaboration et al. 2020).

2. Data and Observations

In this section, we present an overview of the LoTSS DR1 and the HDR4 data sets used in this work. This includes the value-added catalogs from previous efforts, the matching methodology, and the spectral extractions.

2.1. LOFAR Two-metre Sky Survey

The LoTSS is a high-resolution 120–168 MHz survey centered at 150 MHz, with a median sensitivity of $S_{150\,\mathrm{MHz}} = 71\mu$Jy beam^{-1} and a point-source completeness of 90% at an integrated flux density of 0.45 mJy (T. W. Shimwell et al. 2017). The spatial resolution of the images is 6″ and the astrometric accuracy of the data is within 0.″2 (T. W. Shimwell et al. 2019). The first data release includes 424 deg^2 in the HETDEX Spring Field (R.A.s between 10^h45^m and 15^h30^m and declinations ranging from 45°00′ to 57°00′). There are 325,694 sources in the first data release for the HETDEX Spring Field. Additionally, the LoTSS DR1 provides the astrometric precision needed to identify optical and infrared counterparts. We use the first data release because it specifically targeted the HETDEX Spring Field, and while LoTSS DR2 (formed by two regions centered at R.A. $= 12^h45^m00^s$, decl. $= +44°30'00''$and R.A. $= 1^h00^m00^s$, decl. $= +28°00'00''$, spanning 4178 and 1457 deg^2, respectively) is available, it only expands upon the area of the first data release, essentially making DR1 and DR2 interchangeable for our purposes.

W. L. Williams et al. (2019) combined Pan-STARRS $grizy_{\mathrm{P1}}$ photometry (K. C. Chambers et al. 2016) with data from the Wide-field Infrared Survey Explorer (WISE; E. L. Wright et al. 2010) over the LoTSS DR1 region. Using a combination of statistical techniques and visual identification, W. L. Williams et al. (2019) constructed a color- and magnitude-dependent likelihood ratio method for statistical identification. This resulted in a value-added catalog[12] with 318,520 radio sources, of which 231,716 (73%) have optical and/or IR identifications in Pan-STARRS and WISE.

Along with optical/IR counterparts for each radio source, the LoTSS DR1 value-added catalog includes photometric redshift estimates from K. J. Duncan et al. (2019). These

[11] VIRUS is a joint project of the University of Texas at Austin, Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University (TAMU), Max-Planck-Institut für Extraterrestriche-Physik (MPE), Ludwig-Maximilians-Universität München, Pennsylvania State University, Institut für Astrophysik Göttingen, University of Oxford, Max-Planck-Institut für Astrophysik (MPA), and The University of Tokyo.

[12] https://lofar-surveys.org/public/LOFAR_HBA_T1_DR1_merge_ID_optical_f_v1.2b_restframe.fits

estimates are crucial for identifying properties of the radio sources, as the Sloan Digital Sky Survey provides spectroscopic redshifts for less than 1 percent (2690) of the sources. In the near future, the William Herschel Telescope Enhanced Area Velocity Explorer (WEAVE; G. Dalton et al. 2012, 2014) multi-object and integral field spectrograph will measure redshifts of over a million LoTSS sources as part of the WEAVE-LOFAR survey (D. J. B. Smith et al. 2016). We can take the first step in informing that large effort by increasing the known spectroscopic redshifts by combining the LoTSS DR1 value-added catalog with the fourth data release from the HETDEX survey.

2.2. HETDEX Data Release 4

The HETDEX survey is designed to measure the Hubble expansion parameter and angular diameter distances by using the spatial distribution of nearly one million LAEs. The survey employs the VIRUS instrument, which is composed of a set of 78 fiber integral field units (IFUs) feeding 156 identical spectrographs that produce 34,944 spectra covering the wavelength range 3470–5540 Å at a resolving power $R \approx 800$. The IFUs are arrayed in a grid pattern on the sky with a fill factor that is $\simeq 1/4.5$, covering 56 arcmin^2 within an $18'$ diameter field. Each IFU covers a solid angle of approximately $51'' \times 51''$ and feeds two spectrographs, each with 224 fibers. The individual fibers are $1\farcs5$ in diameter, and the spacing between the fiber centers is $2\farcs2$. During HETDEX observations, a dither pattern of three exposures nearly fills these gaps (∼94% sky coverage; see G. J. Hill et al. 2021 for details). The astrometric accuracy of the fiber positions for HETDEX is $0\farcs35$ (K. Gebhardt et al. 2021).

The HETDEX survey serves as the primary observing mode at the HET during dark-sky conditions, accumulating a wealth of data. The internal fourth data release completed observations on 2023 August 31, and it includes ∼67.48 deg^2 of fiber sky coverage with exposures from 2017 August through 2023 August. The majority of these observations are in the HETDEX Spring Field with 67.48 deg^2 sky coverage and 523 million fiber spectra. The data reductions are described in K. Gebhardt et al. (2021), but to summarize the HETDEX team produced two main products—a full set of flux-calibrated fiber spectra and a catalog of automatically detected and classified sources (E. Mentuch Cooper et al. 2023).

To construct the source catalog the HETDEX team ran two object detection algorithms: one designed to find emission-line sources and the other built to search for continuum emission (E. Mentuch Cooper et al. 2023). From these two raw catalogs, source sizes were defined using a friends-of-friends algorithm to avoid multiple detections of the same object. The team then took a multipronged approach to source classification and redshift assignment. The details of the classification and redshift assignment can be found in E. Mentuch Cooper et al. (2023). In short, each source was classified as a star (STAR), a low-redshift galaxy with no [O II] emission (LZG), an [O II] emitting galaxy (O II), an LAE, or an AGN.

The HDR4 catalog is dominated by emission-line galaxies and includes 920,715 LAE candidates with $1.88 < z < 3.52$ and with a signal-to-noise ratio (S/N) greater than 4.8. Also included in the catalog are 451,224 [O II] emitting galaxies at $z < 0.5$, 775,063 stars, 98,801 low-redshift ($z < 0.5$) galaxies without emission lines, and 48,194 AGN. The catalog provides sky coordinates, redshifts, line identifications, classification information, line fluxes, [O II] and Lyα line luminosities when applicable, and spectra for all identified sources processed by the HETDEX detection pipeline.

Although the catalog provides many of the products that we need, we can supplement the HETDEX effort by extracting spectra at the precise locations of the LoTSS sources and then running similar classification tools for a complete redshift analysis of the combined catalogs.

2.2.1. HETDEX Spectral Extractions

We use the HETDEX-API (E. Mentuch Cooper et al. 2023) to extract a spectrum for the 28,705 sources in the LoTSS DR1 catalog with fiber coverage in HDR4.[13] For the 28,705 sources, we first collect all fiber spectra within a $3\farcs5$ radius. Using the seeing measured from the VIRUS data (see K. Gebhardt et al. 2021 for details), we construct a Moffat point-spread function (PSF) model ($\beta = 3.5$; A. F. J. Moffat 1969). At each wavelength, we shift our fiber locations following the differential atmospheric refraction models for the fixed-altitude HET and convolve the PSF with the VIRUS fibers. This calculates the fraction of the object's light covered by each fiber. Using these fiber coverage values as weights, we normalize the weights to 1, retaining the normalization value, and perform a weighted extraction using the K. Horne (1986) optimal extraction formula. Finally, the resultant spectrum is corrected to a total flux using the normalization value. Within a $3\farcs5$ aperture, the total fiber coverage is between 90% and 95%.

Although a PSF spectral extraction is not a natural method for each source in the catalog, it provides a higher S/N methodology than simpler aperture extractions and only introduces a minimal chromatic flux response issue for extended sources. We are not immediately concerned with the absolute calibration of our spectral extractions, as the primary goal is the determination of redshift. A more appropriate extraction can be done for individual science cases starting from the information driven by the description below.

The vast majority of our sample have an average continuum S/N of less than 2 per 2 Å pixel in the wavelength window of 4670–4870 Å. However, it is clear that if the S/N is greater than ∼4, a clear source classification and redshift measurement are possible by eye. Although the sample is not too large for visual analysis, automatic tools with repeatable and similar success rates are available for this purpose (e.g., A. S. Bolton et al. 2012).

3. Classifications and Redshifts

The methodology for our classification scheme begins with the spectral extractions at the sky positions of each LoTSS DR1 source. We use Diagnose (M. Debski & G. Zeimann 2024), a spectral classification code developed for the HETVIPS catalog (G. R. Zeimann et al. 2024), to automatically classify each source and assign a redshift.

3.1. Diagnose

The Diagnose code assigns one of four classifications for each source (star, galaxy, quasar, or unknown) while returning a redshift estimate for the galaxies and quasars and a velocity estimate for the stars. Diagnose determines a spectral classification and redshift estimate for each source via a χ^2

[13] https://github.com/HETDEX/hetdex_api

minimization for linear combinations of principal component templates. In particular, Diagnose uses a principal component analysis with the templates of redrock,[14] which include 10 components for galaxies and 4 components for quasars. Stars are classified by type, with six components for B, A, F, G, K, M, and white dwarfs.

By convolving the high-resolution templates of redrock to the $R \approx 800$ resolution of VIRUS and then fitting these templates to the data using a range of redshifts, Diagnose computes three best-fit χ^2 values: one for stellar type and velocity, one for galaxy type and redshift, and one for a quasar and redshift. Diagnose then compares the best fit of these three χ^2_ν values to the second-best fit and evaluates the difference against a statistical threshold. If this difference is larger than a statistical threshold, the source is classified as the best-fit template (i.e., star, galaxy, or quasar). If the difference is not larger than the threshold, the source is classified as unknown. Using Diagnose, the HETDEX-LOFAR Spectroscopic Redshift Catalog is able to produce classifications and redshift estimates for sources within LoTSS with no known spectroscopic redshifts.

The power of Diagnose is in identifying strong features, usually in the continuum. As the sources become fainter with lower S/N in the continuum, Diagnose becomes less effective and often returns an unknown label. However, many of the radio sources in the HETDEX-LOFAR catalog have strong emission features but weak continua. For these sources, we can use the HDR4 catalog to both check our initial Diagnose classifications and supplement our identifications.

3.2. LoTSS-HDR4 Catalog Matching

To match the LoTSS data set to sources in the HDR4 catalog, we use the HETDEX-API, including the ElixerWidget (D. Davis et al. 2023). We began by using an initial matching radius of 5″ to find counterparts for the 28,705 LoTSS sources. Given the astrometric uncertainty of the two catalogs, this should be a more than ample starting threshold. We found 7409 matches and plot the ΔR.A. versus Δdecl. in Figure 1. We fit a 2D uniform + Gaussian distribution model in ΔR.A. and Δdecl. space to determine the matching standard deviation and estimate the spurious match fraction. We found a standard deviation of 0.″61 in ΔR.A. and 0.″65 in Δdecl. Thus, we use a final matching radius of 2″, which is roughly 3 times the standard deviation. For a matching radius of 2″ we estimate that only ∼5% of the counterparts are spurious by subtracting the area under the Gaussian distribution model from the area under the 2D uniform distribution model and dividing by the area under the 2D uniform distribution model.

3.3. Combining Diagnose and HDR4

Here, we describe how we combine our Diagnose classifications/redshifts with the classifications/redshifts obtained from the LoTSS-HETDEX catalog matching. Starting from the 28,705 radio sources, Diagnose confidently identified 6480 objects as a star, galaxy, or quasar; the remaining 21,081 sources did not have a reliable classification. Of those, 998 objects had insufficient spectral coverage for either a Diagnose or a catalog label; this is due to masking of bad fibers/amplifiers in the data set after the initial spatial matching.

[14] https://github.com/desihub/redrock-templates

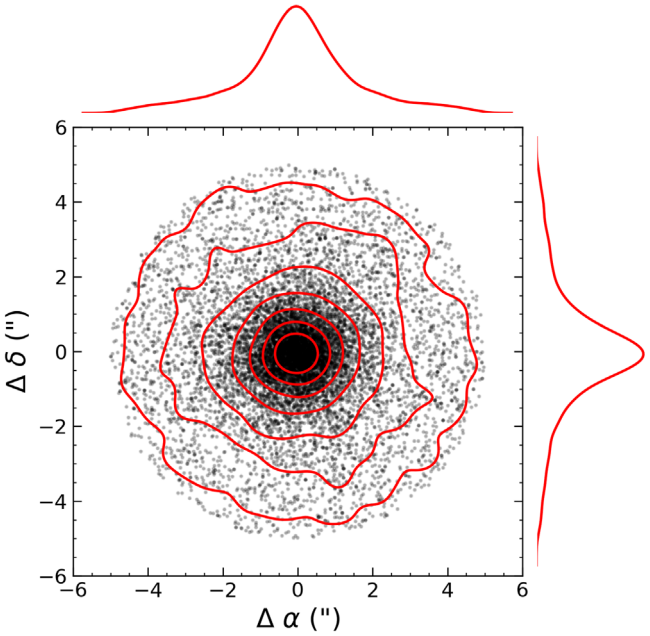

Figure 1. Comparison between $\Delta\alpha$ and $\Delta\delta$ at initial matching radius of 5″ for the 7409 matches found between the HETDEX HDR4 catalog and the LoTSS sources. The x-axis represents the $\Delta\alpha$ and the y-axis represents the $\Delta\delta$. Sources within a radius of 2″ have a less than 5.43% chance of being false matches when fitted with a uniform + Gaussian distribution model.

For the 7409 spatial matches within the LoTSS-HETDEX catalog, we collected the classification and redshift from HDR4. When comparing the sources with both a Diagnose and HDR4 redshift, we find good agreement, with 92.3% of the objects agreeing to within $\Delta z = 0.05$. This is not entirely surprising, as the HDR4 classification scheme uses Diagnose for sources with continuum g-band magnitudes brighter than 22. For the objects with discrepant redshifts, the most common reason was Diagnose labeling an emission line as [O II], rather than Lyα. This occurred 3.03% of the time. As noted before, we also expect a ∼5% spurious match fraction between HDR4 and LoTSS, which may account for the remaining disagreement between the two redshift estimates.

Comparing the HDR4 spectroscopic redshifts with our Diagnose redshifts, we find 4908 sources in common. Figure 2 compares the two redshifts; the outlier fraction is 6.7%, while the standard deviation of the remaining objects is $\sigma_z = 0.0001$. Figure 2 also highlights groups of particular interest within the set of objects with discrepant redshifts. For each of these groups, we investigated the spectra by eye and determined criteria for determining which redshift is the best fit. These criteria and the process for choosing the best redshift are detailed in the Appendix. After applying all of the criteria to the different groups of spurious matches, the outlier fraction reduces from 6.7% to 2.3%.

3.4. LoTSS DR1 Value-added Catalog Redshifts

Within 28,705 LoTSS sources, there are 11,807 objects with photometric redshifts and 2690 sources with spectroscopic redshifts in the value-added catalog from W. L. Williams et al. (2019) and K. J. Duncan et al. (2019). The majority of spectroscopic redshifts were compiled from the Sloan Digital Sky Survey Data Release 14 (DR14; B. Abolfathi et al. 2018). These redshifts were supplemented by additional spectroscopic data from a range of deep optical surveys in the literature,

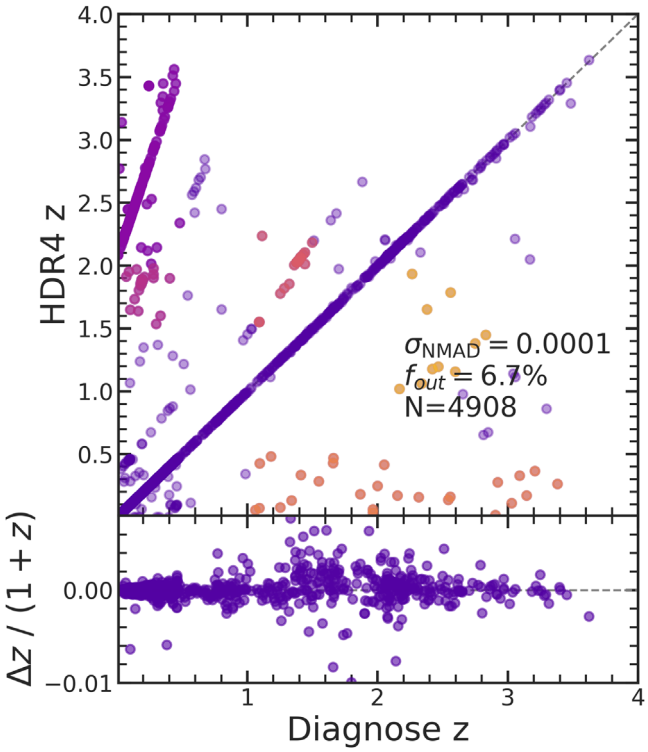

Figure 2. Comparison between sources with both HDR4 redshifts and Diagnose redshifts. There is good agreement between the two redshifts, though there is some scatter. The different colored groups represent areas of interest that were further investigated (see Section 3.3) to make the correct redshift assignments. The normalized median absolute deviation (NMAD) and outlier fraction are calculated as in I. G. Momcheva et al. (2016).

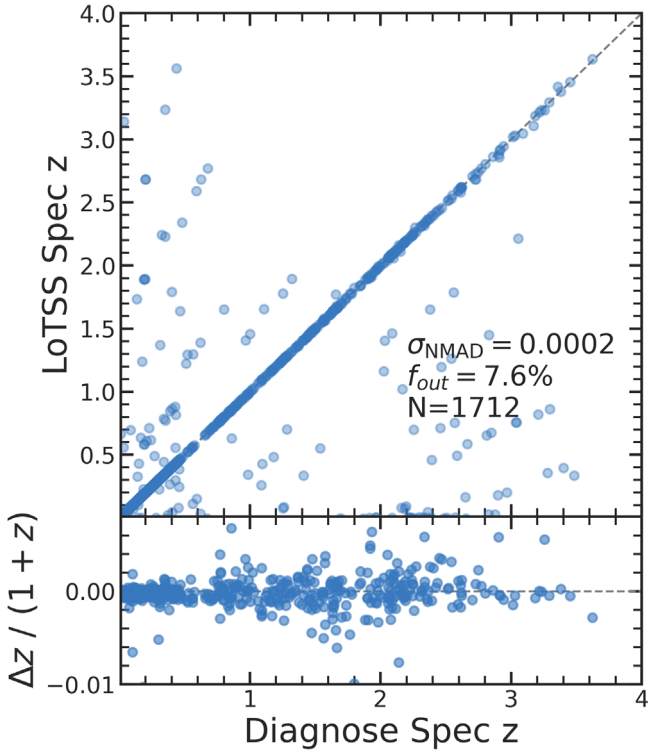

Figure 3. Comparison between LoTSS previously determined spectroscopic redshifts and Diagnose spectroscopic redshifts for the same sources. In total, there were 1098 LoTSS sources with previous spectroscopic redshift counterparts. The normalized median absolute deviation (NMAD) and outlier fraction are calculated as in I. G. Momcheva et al. (2016). The LoTSS redshifts and Diagnose redshifts are in good agreement.

mostly covering the Extended Groth Strip within the HETDEX Spring Field. We refer to the spectroscopic redshifts in the value-added catalog as archival redshifts.

Comparing the archival spectroscopic redshifts with our HETDEX-LOFAR catalog, we find 1701 sources in common. The vast majority (∼75%) of the archival redshifts that are not in our catalog are in the redshift desert of VIRUS ($0.5 < z < 1.9$) where there are no strong emission lines. Investigating the overlapping sources, we find good agreement between our spectroscopic redshifts and those in the literature. Figure 3 shows the outlier fraction is 7.6% with a standard deviation of nonoutlying sources $\sigma_z = 0.0002$.

We also compared our spectroscopic redshifts to the photometric estimates from the LoTSS value-added catalog. Figure 4 shows this comparison over the range $0.0 < z < 0.5$; we find a good agreement between the two measurements with $\sigma_z = 0.0614$.

3.5. Classification and Redshift Methodology

The HETDEX-LOFAR spectroscopic redshifts and classifications have three origins: Diagnose, HDR4, and archival. We discussed the combination of Diagnose and HDR4 redshifts in Section 3.3, and with all three origins we follow a similar logic. If Diagnose has a robust redshift, this is used. If not and HDR4 has a redshift, then this is used. Finally, if neither Diagnose nor HDR4 provides a redshift for the source but the value-added catalog does, then we use the archival spectroscopic redshift. We prioritize archival redshifts last in order to maximize the number of new redshifts determined for this catalog. Figure 5 shows the final redshift distribution for the 9710 sources and the origin of the redshift. The overall detection and classification pipeline is outlined in Figure 6.

Our classification follows mostly from the redshift of the source and the origin catalog. We group all sources labeled "STAR" by Diagnose or HDR4 together as "STAR." We group all objects labeled "QSO" or "AGN" in Diagnose or HDR4, respectively, as "AGN"; this is done for all redshifts $0.0 < z < 3.5$. Classifications of "LZG" and "O II" from HDR4 and "GALAXY" from Diagnose are all grouped under the label "LOWZGAL" for galaxies $0.0 < z < 0.5$. Classifications of "LAE" from HDR4 are grouped as "HIGHZGAL" for systems with $1.9 < z < 3.5$. Finally, if the redshift comes from the archive, we label the group "ARCHIVE" with $0.0 < z < 3.5$. So, our final five labels are "STAR," "AGN," "LOWZGAL," "HIGHZGAL," and "ARCHIVE." Table 1 includes a breakdown of the number of sources at different steps of the classification process, including the final catalog size.

3.6. HETDEX-LOFAR Spectroscopic Catalog

Our final compiled spectroscopic redshift catalog includes 9710 total redshifts: 197 "STAR," 804 "AGN," 6394 "LOWZGAL," 1075 "HIGHZGAL," and 757 "ARCHIVE." Table 2 explains the column names for the final catalog. In Figure 7, we show three example spectra for each of the five labels.

4. Star Formation at Radio Wavelengths

4.1. Spectral Energy Distribution Fitting

We limited our analysis of HETDEX-LOFAR galaxies to the 6499 objects with $0.01 < z < 0.47$ to ensure [O II] was in the

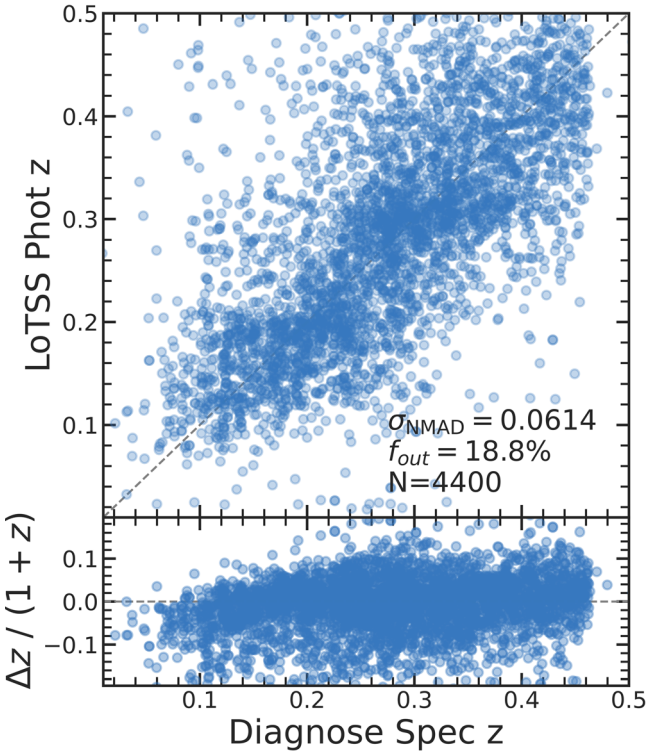

Figure 4. Comparison between LoTSS photometric redshifts and Diagnose spectroscopic redshifts for the same sources. In total, there were 4400 LoTSS sources with photometric redshifts and no previous spectroscopic redshift counterparts. The vertical bound created by the data points is a result of the cutoff at the highest possible redshift at VIRUS wavelengths for an [O II] emitter. The normalized median absolute deviation (NMAD) and outlier fraction are calculated as in I. G. Momcheva et al. (2016). This relationship is to be expected for a photo-z to spec-z comparison.

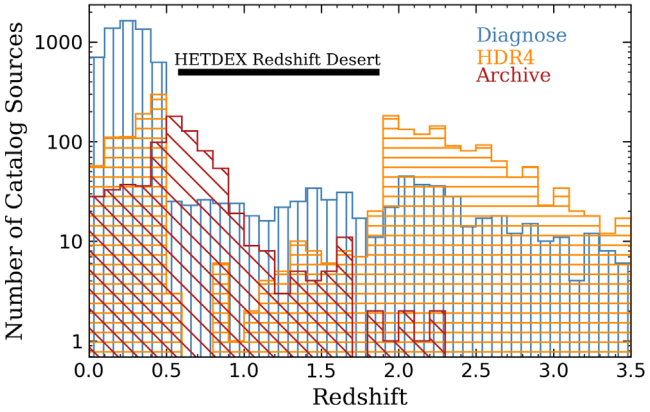

Figure 5. We show the distribution of redshifts in the HETDEX-LOFAR spectroscopic catalog. There are 9710 sources in the catalog, the majority of which are between $0.0 < z < 0.5$ and $1.9 < z < 3.5$. There is a void of redshifts between those two regions due to a lack of strong features present in the VIRUS wavelength bandpass. The stacked histogram in blue shows the redshifts from Diagnose, which make up the majority of the sample and are mostly lower redshift objects. The stacked histogram in orange shows the redshifts from HDR4; these are more evenly distributed between low and high redshift. Finally, the stacked histogram in red are the redshifts from the archive not in the other two classifiers.

VIRUS bandpass. For each galaxy, we collected the Pan-STARRS *grizy*$_{\rm P1}$ and WISE W1W2 photometry in the W. L. Williams et al. (2019) catalog, and then further restricted our sample to those sources with a Pan-STARRS $g_{\rm P1}$-band detection; this reduced our sample to 5919 systems. Although this photometry alone is often enough to derive quantities such as SFR and stellar mass from spectral energy distributions (SEDs), we can also utilize our VIRUS spectroscopy to further inform the fitting. For consistency, we normalized our VIRUS spectra to the Pan-STARRS $g_{\rm P1}$-band photometry, and then calculated 10 synthetic narrowband values in chunks of 200 Å across the bandpass. We also used `ppxf` (M. Cappellari 2023) to quickly model the underlying stellar continuum to measure the [O II] emission. Using the 7 bands of photometry, 10 synthetic narrow bands of the VIRUS spectroscopy, and [O II] emission, we estimated the stellar masses, SFRs, and dust attenuation via SED fitting using `MCSED` (W. P. Bowman et al. 2020).

`MCSED` is a flexible SED-fitting code that allows users to supply both photometry and emission-line fluxes to fit the stellar populations of a galaxy. `MCSED` implements a stellar library generated by the Flexible Stellar Population Synthesis code (C. Conroy et al. 2009; C. Conroy & J. E. Gunn 2010) employing PADOVA isochrones (G. Bertelli et al. 1994; L. Girardi et al. 2000; P. Marigo et al. 2008), a self-consistent prescription for nebular line and continuum emission given by the grid of CLOUDY models (G. J. Ferland et al. 1998, 2013) generated by N. Byler et al. (2017), and a G. Chabrier (2003) initial mass function (IMF). We adopted an eleven-parameter model, with the variables being stellar metallicity (ranging from ∼1% to 150% solar metallicity), a nonparametric six-age-bin star formation history (using a constant SFR within each bin defined at ages of 0.001, 0.03, 0.1, 0.3, 1.0, 3.6, and 13.2 Gyr), a single-parameter dust attenuation law (D. Calzetti et al. 2000), and a three-parameter dust emission model from B. T. Draine & A. Li (2007) constrained by energy balance between absorption and emission. The nebular metallicity was fixed to the stellar metallicity, and the ionization parameter of the nebular emission was fixed at $\log(U) = -2.5$. Changing the adopted fitting assumptions (especially the SFR history) can systematically affect the stellar masses at the level of ∼0.3 dex (C. Conroy et al. 2009; C. Conroy & J. E. Gunn 2010).

`MCSED` utilizes the `emcee` Python module (D. Foreman-Mackey et al. 2013) with initial positions defined by a random Gaussian ball near the middle of the range of allowed values for each parameter and a small but generous sigma to avoid the initial boundaries yet explore the available phase space. We ran the fitting using a Monte Carlo Markov Chain (MCMC) approach with 40 walkers and 800 steps. Convergence is always a challenge in Monte Carlo methods, and with 11 free parameters, the choice of 40 walkers and 800 steps was a compromise between convergence and computation cost.

4.2. Star Formation Rate and 150 MHz Luminosity

For our sample of galaxies, we used the `MCSED` results to examine the correlation between 150 MHz luminosity and SFR, as well as the secondary stellar mass dependence. Figure 8 shows the results for each of these relationships. All of the galaxies studied have SFR and stellar mass estimates that were derived from energy balance spectral energy distribution fitting using redshifts and aperture-matched forced photometry from the LoTSS Deep Fields data release. The first panel in Figure 8 shows the correlation between SFR and 150 MHz luminosity. There is tight correlation between the two quantities, though there is some scatter caused by the secondary mass dependence acknowledged by D. J. B. Smith et al. (2021). This mass dependence can be seen via the color bar. The

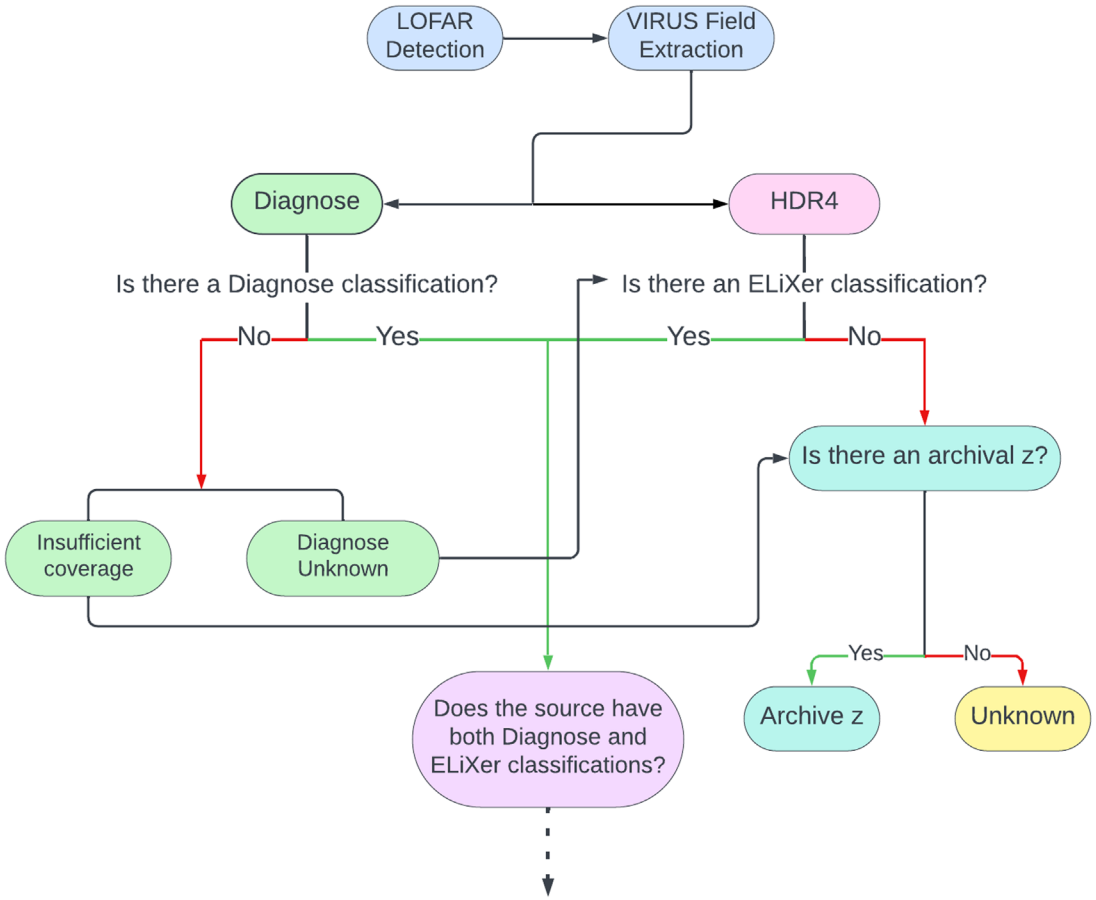

Figure 6. HETDEX-LOFAR redshift and classification pipeline. The process begins with the LOFAR detection, followed by the extraction of HETDEX spectra. Once we have the HETDEX spectra, we run them through Diagnose and ELiXer to determine whether they have a classification and redshift from either source. If they have either a Diagnose or ELiXer redshift, we accept this value; if they have both and the redshifts disagree, we perform a detailed examination for the reason behind the discrepancy. The specifics of this examination are detailed in the Appendix. If neither Diagnose nor ELiXer provides a redshift, we see if there is an archival value. If so, we accept this value. If not, the source remains unclassified.

Table 1
Number of Sources Relevant to Different Steps of the Classification and Redshift Assignment Pipeline Including Final Catalog Size

Number of Sources	Description
325,694	Sources in LoTSS DR1
28,705	Spectral matches between LoTSS DR1 & HETDEX DR4
4908	Sources w/ Diagnose & ELiXer redshifts
9710	Number of spectroscopic redshifts in final HETDEX-LOFAR catalog
9087	New spectroscopic redshifts in final catalog
197	"STAR" in final catalog
804	"AGN" in final catalog
6394	"LOWZGAL" in final catalog
1075	"HIGHZGAL" in final catalog
757	"ARCHIVE" in final catalog

Table 2
Description of the Catalog Columns

Column Name	Description	Data Type
objID	LoTSS Object ID	string
source_name	Source name	string
R.A.	PanSTARRS1 R.A. (J2000)	float
Decl.	PanSTARRS1 decl. (J2000)	float
z_diagnose	Best-fit redshift from Diagnose	float
z_hdr4	Best-fit redshift from ELiXer	float
z_archive	Spectroscopic redshift from value-added LoTSS catalog	float
z_best	HETDEX-LOFAR redshift	float
z_best_src	1 = Diagnose, 2 = HDR4, 3 = Archive	integer
classification	STAR, AGN, LOWZGAL, HIGHZGAL, or ARCHIVE	string
log_mass	MCSED derived stellar mass	float
log_SFR	MCSED derived star formation rate	float
log_L150	MCSED derived 150 MHz luminosity	float

Note. "z best" and "classification" reflect final HETDEX-LOFAR catalog values. (This table is available in its entirety in machine-readable form in the online article.)

second panel demonstrates the correlation between stellar mass and 150 MHz luminosity. There is a strong correlation here, which is to be expected as we anticipate a secondary mass dependence. Additionally, we referenced the D. J. B. Smith et al. (2021) calculation for 150 MHz luminosity to determine the predicted luminosity expected for our MCSED results, which is shown in the third panel. The black dashed line represents a one-to-one correlation. All of our data follow this trend and are tightly correlated. This comparison acts as a check on our derived values for 150 MHz luminosity, stellar mass, and SFR. The majority of our derived values match closely to those

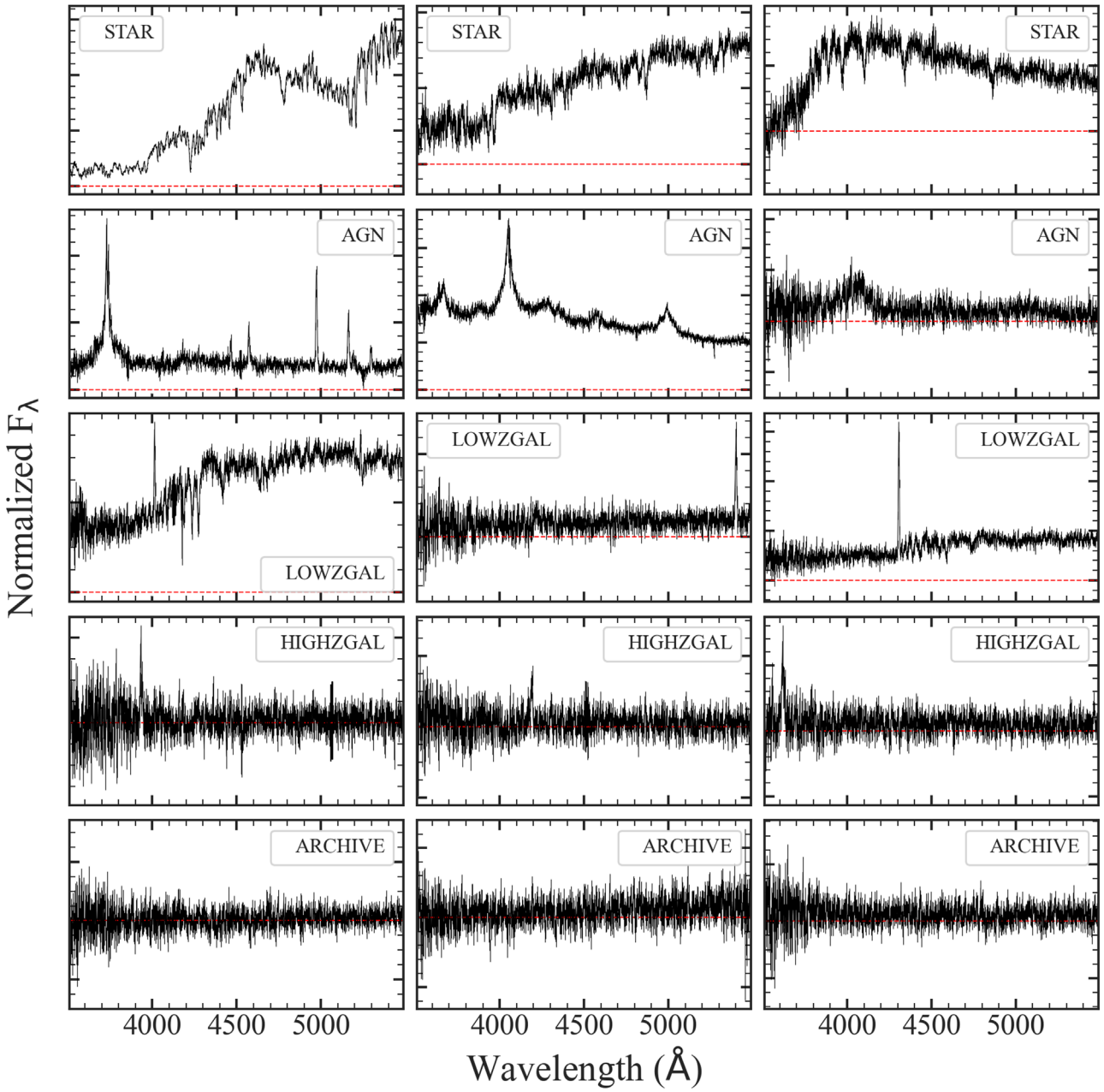

Figure 7. We show 15 example spectra from the HETDEX extractions of LoTSS sky positions: three each with labels of "STAR," "AGN," "LOWZGAL," "HIGHZGAL," and "ARCHIVE." We also mark the null flux density with a dashed red line.

predicted by the Smith relation. Overall, we were able to measure SFR, 150 MHz luminosity, and stellar mass for 6499 galaxies.

4.3. Line Ratio Diagnostics

We used PPXF to measure the line fluxes for our individual galaxies, particularly focusing on the [O III], [O II], and [Ne III] emission lines. We measure these lines in particular due to their use in radio astronomy. [O III] is used to trace ionized outflows from radio sources because of its sensitivity to the impact of radiation and jets (e.g., P. Kukreti et al. 2023). Additionally, these three lines can be useful in line ratios as indicators of ionization parameter. The most commonly used diagnostic of the ionization parameter is [O III]/[O II] (O3O2; e.g., D. Alloin et al. 1978; J. A. Baldwin et al. 1981); however, the wavelength range between [O III] and [O II] makes this line ratio diagnostic radio sensitive to extinction effects. As an alternative, [Ne III]/[O II] (Ne3O2) can act as a similar diagnostic of ionization parameter that is radio insensitive to reddening effects (e.g., E. M. Levesque & M. L. A. Richardson 2014). The similar short wavelengths of [Ne III] and [O II] also make the diagnostic usable at larger redshifts ($z \sim 1.6$) than O3O2 (T. Nagao et al. 2006). The benefits of Ne3O2 as a diagnostic of ionization parameter as compared to those of O3O2 led us to solely examine the Ne3O2 line ratio in our sample. Because many of the measured lines are weak, we also computed the biweight stack of the HETDEX spectra binned by stellar mass, with each spectrum normalized by its median continuum value in the rest-frame wavelength range of $3750\ \text{Å} < \lambda < 3850\ \text{Å}$ for Ne3O2. We also limited our sample to galaxies with $z < 0.4$ to ensure we detect the [Ne III] and [O II] lines.

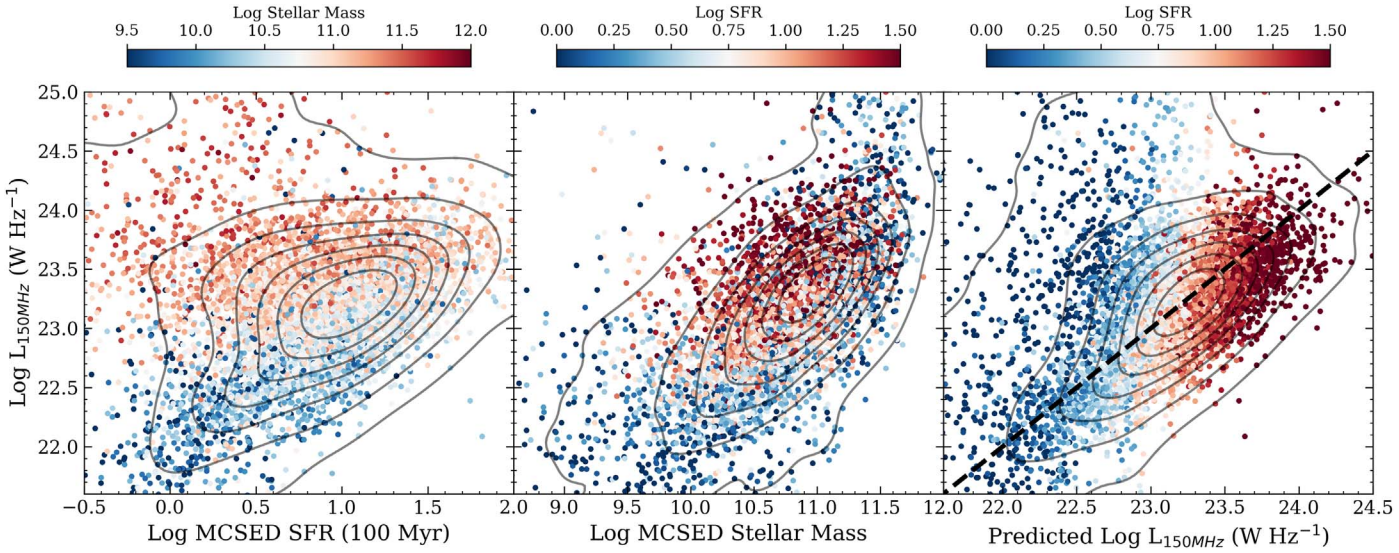

Figure 8. MCSED results for the 6499 galaxy sample with $0.01 < z < 0.47$. The left panel demonstrates a tight correlation between SFR and 150 MHz luminosity with expected scatter from the secondary mass dependence show in D. J. B. Smith et al. (2021). The individual points are colored by stellar mass and show this dependence. The center panel demonstrates the correlation between stellar mass and 150 MHz luminosity, as colored by SFR. The right plot shows the correlation between our derived 150 MHz luminosity and the D. J. B. Smith et al. (2021) predicted 150 MHz luminosity. The black dashed line represents a one-to-one correlation. The individual points are colored by SFR.

The line ratio diagnostics used are indicators of ionization and relationship to AGN activity; however, they are not definitive criteria to determine contribution from AGN activity. To further understand the properties of our data set, we used the relation between SFR and 150 MHz luminosity derived in P. N. Best et al. (2023). This relation can be used to set an N-σ cutoff above the P. N. Best et al. (2023) ridge line, which can help determine which galaxies are star-forming and which are dominated by AGN activity. Galaxies that fall within the N-σ cutoff are star-forming, while those above the cutoff have AGN activity.

We further explored this relation by examining the relationship between stellar mass and the N-σ offset from the P. N. Best et al. (2023) relation (Figure 9). AGN typically have harder ionization fields, meaning that their $\log_{10}$(Ne3O2) line ratios should be greater than 1. By coloring the scatter in Figure 9 by the Ne3O2 ratio, we see that none of the galaxies with $z < 0.4$ have $\log_{10}(\mathrm{Ne3O2}) > 1$; however, as galaxies reach $\log_{10}(M/M_\odot) > 10.5$, Ne302 increases with the offset from the P. N. Best et al. (2023) relationship. This suggests that there could be a greater contribution from AGN.

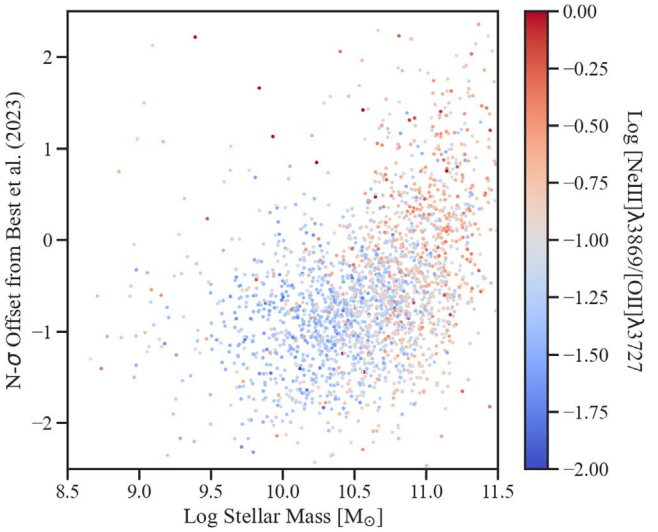

Figure 9. Stellar mass vs. N-σ offset from the P. N. Best et al. (2023) relation between SFR and 150 MHz luminosity colored by $\log_{10}$(Ne3O2) line ratio. AGN are anticipated to have $\log_{10}(\mathrm{Ne3O2}) > 1$, but none of our sources exhibit such ionization hardness. At $\log_{10}(M/M_\odot) > 10.5$, there is an increase in offset with Ne3O2, indicating that there could be more AGN contribution for these sources.

4.4. The SFR–150 MHz Luminosity Relation

We determined the SFR–$L_{150\mathrm{MHz}}$ relation as follows. First, we applied a mass cut at $\log_{10}(M/M_\odot) < 11.0$. This mass cut arises from the analysis of Figure 9, which shows increasing AGN contribution above $\log_{10}(M/M_\odot) \approx 10.5$. To ensure that our fit was based mostly on star-forming galaxies, we removed the population with potential AGN contribution from our sample used for fitting. For the remaining sources, we calculated the errors on SFR, stellar mass, and $L_{150\mathrm{MHz}}$ and found an average $\log_{10}$ error of ± 0.216 [$M_\odot\,\mathrm{yr}^{-1}$] for SFR, ± 0.132 [$M_\odot$] for stellar mass, and ± 0.070 [$\mathrm{W\,Hz}^{-1}$] for $L_{150\mathrm{MHz}}$.

To determine best-fit parameters, we adopted the form of the mass-dependent SFR–$L_{150\mathrm{MHz}}$ relationship from G. Gürkan et al. (2018):

$$L_{150\mathrm{MHz}} = L_C \psi^{\beta} \left(\frac{M_*}{10^{10} M_\odot} \right)^{\gamma} \quad (1)$$

where L_C is the 150 MHz luminosity of a galaxy with $M_* = 10^{10} M_\odot$ and $\psi = 1\ M_\odot\,\mathrm{yr}^{-1}$. We find the best-fit values of $\log_{10} L_C = 22.341 \pm 0.016$, $\beta = 0.526 \pm 0.017$, and $\gamma = 0.384 \pm 0.017$, determined using the `emcee` (D. Foreman-Mackey et al. 2019) MCMC algorithm with 15 walkers and a chain length of 10,000 samples. Our best-fit SFR–$L_{150\mathrm{MHz}}$ relation is, therefore, $\log_{10} L_{150\mathrm{MHz}} = (22.341 \pm 0.016) + (0.526 \pm 0.017)\ \log_{10}(\psi/M_\odot\,\mathrm{yr}^{-1}) + (0.384 \pm 0.017)\ \log_{10}(M/10^{10} M_\odot)$.

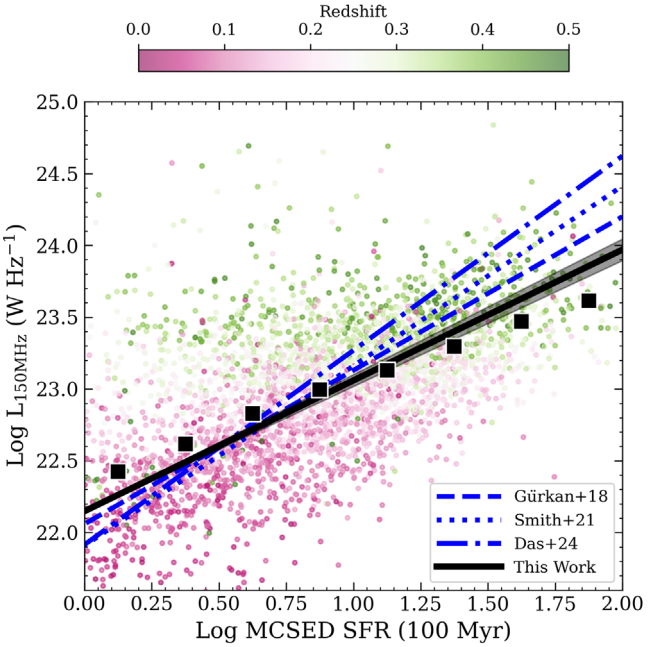

Figure 10. Distribution of derived radio luminosity vs. MCSED star formation rate (100 Myr) for the HETDEX-LOFAR galaxies with $0.01 < z < 0.47$. The points represent the portion of the sample with $\log_{10}(M/M_{\odot}) < 11.0$ and are colored by redshift. The black squares represent the biweight of $\log_{10}(L_{150MHz})$ for bins of $\log_{10}(\psi)$. Our derived relationship for radio luminosity vs. star formation rate is shown by the black solid line. Different fits from literature are shown with blue lines (dashed—G. Gürkan et al. 2018; dotted—D. J. B. Smith et al. 2021; dashed–dotted—S. Das et al. 2024).

We compared this mass-dependent line fit to those found by G. Gürkan et al. (2018), D. J. B. Smith et al. (2021), and S. Das et al. (2024) in Figure 10. G. Gürkan et al. (2018) obtained best-fit estimates of $\log_{10}L_C = 22.13 \pm 0.01$, $\beta = 0.77 \pm 0.01$, and $\gamma = 0.43 \pm 0.01$. D. J. B. Smith et al. (2021) found best-fit estimates of $\log_{10}L_C = 22.218 \pm 0.016$, $\beta = 0.903 \pm 0.012$, and $\gamma = 0.332 \pm 0.037$. S. Das et al. (2024) obtained best-fit estimates of $\log_{10}L_C = 22.083 \pm 0.004$, $\beta = 0.778 \pm 0.004$, and $\gamma = 0.334 \pm 0.006$. Our line has a shallower slope than the three previously derived mass-dependent expressions. This difference could be a result of several factors. Our relation is derived using spectroscopic redshifts with small error, creating an upper radio luminosity limit that is quite sharp (shown by the coloring of data points in Figure 10). This upper limit could be depressing the steepness of our fit because it is creating an upper limit on radio luminosity. There is also a lower limit caused by the flux density limit of our sample. The combination of both the upper and lower limits on luminosity could result in a lack of objects that would tend to populate the lower left and upper right of Figure 10. The slope of the fit is most influenced by objects at the extremes of radio luminosity and SFR, so if there is limit-related bias present, the slope will also be biased. Additionally, the applied mass cut could have removed a number of the high-luminosity sources, which would also create bias in the slope. The larger sample that will be available with the release of the full HETDEX survey will allow a more thorough analysis of the possible systematic effects in fitting the SFR relation, and this analysis can be investigated further in a future paper.

5. Summary

Combining data from an optical spectroscopic survey, the fourth data release of the HETDEX catalog, and a radio survey, the first data release of the Low Frequency Array LoTSS, we were able to determine intrinsic properties for radio sources present in both fields of view. Using positions in LoTSS we extracted 18,267 spectra from the HETDEX database. We used a robust and automatic classification code called Diagnose, developed by our group for the HET VIRUS Parallel Survey, to determine redshifts and object classifications from the optical spectra. We also use the HETDEX ElixerWidget to source redshifts and object classifications. To determine redshifts and classifications for the remaining unknown spectra, we matched these sources (2″ radius) to object positions in the HETDEX data release 4 catalog. Using these methods, we created the HETDEX-LOFAR Spectroscopic Redshift Catalog with 9710 total redshift values. We group all sources labeled "STAR" by Diagnose or HDR4 together as "STAR." We group all objects labeled "QSO" or "AGN" in Diagnose or HDR4, respectively, as "AGN"; this is done for all redshifts $0.0 < z < 3.5$. Classifications of "LZG" and "O II" from HDR4 and "GALAXY" from Diagnose are all grouped under the label "LOWZGAL" for galaxies $0.0 < z < 0.5$. Classifications of "LAE" from HDR4 are grouped as "HIGHZGAL" for systems with $1.9 < z < 3.5$. Finally, if the redshift comes from the archive, we label the group "ARCHIVE" with $0.0 < z < 3.5$. So, our final five labels are "STAR," "AGN," "LOWZGAL," "HIGHZGAL," and "ARCHIVE."

The compiled catalog includes 197 "STAR," 804 "AGN," 6394 "LOWZGAL," 1075 "HIGHZGAL," and 757 "ARCHIVE" sources.

The focus of this project is assigning redshifts for sources, which, for extragalactic objects, allows one to determine distances and hence many intrinsic properties, most importantly the luminosity. Using line ratio diagnostics such as [Ne III]/[O II], we probed the ionization parameter of the gas. These properties enable AGN excitation to be detected within the star-forming galaxies.

The HETDEX-LOFAR Spectroscopic Redshift Catalog contains the highest substantial fraction of LOFAR galaxies with spectroscopic redshift information and coverage. The catalog also offers an improvement over the archival spectroscopic redshift estimates provided in the LoTSS value-added catalog. It also enables the investigation of SFR tracers with high-quality data, though the results may not be definitive. We derive the SFR, stellar masses, 150 MHz luminosity, and emission lines for ∼75% of our sources with $z < 0.4$, as well as fit a new SFR–L_{150MHz} relationship. Understanding the relationship between radio luminosities and SFR is increasingly important in the upcoming era of SKA, and this work acts as the first step to help inform future work in the radio community such as investigating the connection between ionized outflows traced by [O III] and radio emission, refining radio luminosity functions, and comparison of spectra with resolved subgalactic radio emission. All of the values derived through this work can act as a reference point for adjusting star formation surveys.

The HETDEX-LOFAR Spectroscopic Redshift Catalog provides key physical properties of 9710 objects including R. A., decl., spectroscopic redshift, classification, stellar mass, SFR, and 150 MHz luminosity. These properties will serve to enable science investigations in the radio astronomy community.

5.1. Data Release

The fourth internal data release of HETDEX covers ∼90% of the total survey, making this the first paper in a series about combining HETDEX and LoTSS. By the time HETDEX finishes, we anticipate a final HETDEX-LOFAR sample of ∼40,000 galaxies. This paper includes the release of derived spectroscopic redshifts, classifications, stellar masses, SFR, 150 MHz luminosity, and line fluxes for each source, as well as the spectra. A copy of the HETDEX-LOFAR Spectroscopic Redshift Catalog is available on Zenodo at doi:10.5281/zenodo.14194635. This Zenodo deposit includes a FITS file with the spectra for all 28,705 sources, as well as the derived redshifts, classifications, and MCSED quantities for each source (columns described in Table 2). The deposit also includes the required statements and papers to reference to acknowledge use of data from the HETDEX survey.

Acknowledgments

HETDEX is led by the University of Texas at Austin McDonald Observatory and Department of Astronomy with participation from the Ludwig-Maximilians-Universität München, Max-Planck-Institut für Extraterrestrische Physik (MPE), Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University, Pennsylvania State University, Institut für Astrophysik Göttingen, The University of Oxford, Max-Planck-Institut für Astrophysik (MPA), The University of Tokyo and Missouri University of Science and Technology.

Observations for HETDEX were obtained with the Hobby–Eberly Telescope (HET), which is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximillians-Universität München, and Georg-August-Universität, Göttingen. The HET is named in honor of its principal benefactors, William P. Hobby and Robert E. Eberly. We thank the staff at McDonald Observatory for making this project possible. The Visible Integral-field Replicable Unit Spectrograph (VIRUS) was used for HETDEX observations. VIRUS is a joint project of the University of Texas at Austin, Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University, Max-Planck-Institut für Extraterrestrische Physik (MPE), Ludwig-Maximilians-Universität München, Pennsylvania State University, Institut für Astrophysik Göttingen, University of Oxford, and the Max-Planck-Institut für Astrophysik (MPA).

Funding for HETDEX has been provided by the partner institutions, the National Science Foundation, the State of Texas, the US Air Force, and by generous support from private individuals and foundations.

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 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.

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computing resources that have contributed to the research results reported within this paper: http://www.tacc.utexas.edu.

The Institute for Gravitation and the Cosmos is supported by the Eberly College of Science and the Office of the Senior Vice President for Research at the Pennsylvania State University.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

We also thank the Erickson Discovery Grant for providing support to complete this project during summer 2022. Funding from this grant paved the path to complete the majority of analysis required for this project.

Appendix
Diagnose and HDR4 Redshift

There were 149 sources with $2.0 < z_{\rm HDR4} < 3.5$ and $0.0 < z_{\rm Diagnose} < 0.5$, which demonstrates the common issue of Diagnose identifying an emission line as [O II] while HDR4 identifies the line as Lyα. The classification scheme for these sources is shown in Figure 11. In order to distinguish which is the best-fit redshift for this group, we utilized the HDR4 catalog output "plya_classification," which represents the likelihood that the detected line is Lyα. This criterion ranks each object from 0 to 1 with 1 being a high probability that the line is Lyα. In order to utilize this probability, we determined a cutoff of "plya_classification" = 0.85 using the HDR4

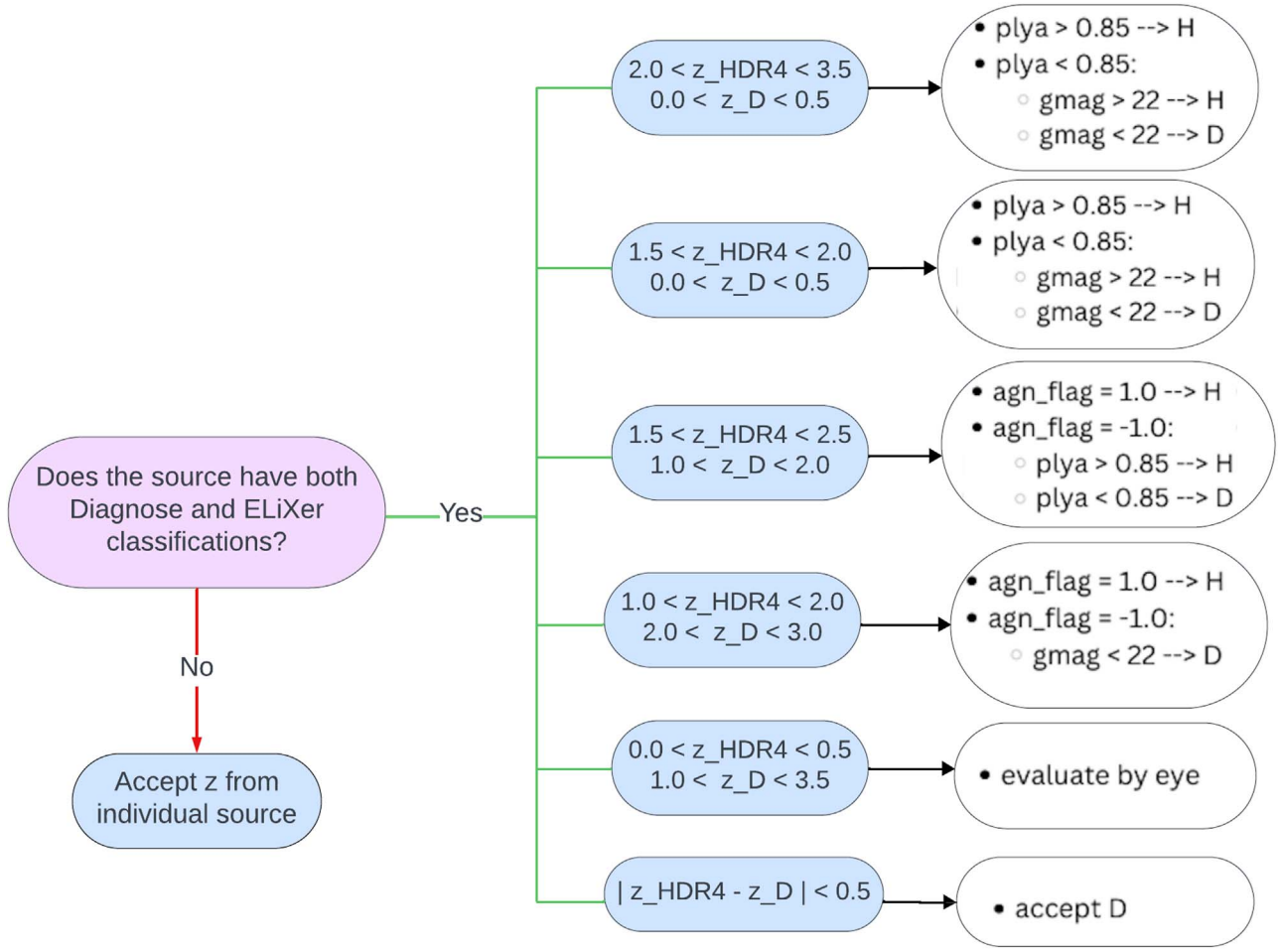

Figure 11. HETDEX-LOFAR redshift and classification pipeline. The process begins with the LOFAR detection, followed by the extraction of HETDEX spectra. Once we have all spectra detected by HETDEX in the LoTSS field, we then run each spectrum through Diagnose and ELiXer to determine whether they have a classification and redshift from either source. If they have either a Diagnose or ELiXer redshift, we accept this value. If they have both, we follow a detailed examination of the difference in redshift between the two sources. The specifics of this examination are detailed in the Appendix. If neither Diagnose nor ELiXer provide a redshift, we see if there is an archival value. If so, we accept this value. If not, the source remains unclassified.

classification scheme that uses Diagnose redshifts for g-band magnitudes brighter than 22. Figure 12 demonstrates how we determined this cutoff using false-positive and false-negative rates for different cutoffs.

To determine the rates, we examined each source's "plya_classifcation" and g-band magnitude at incremental cutoffs between 0 and 1. If the "plya_classification" was greater than the cutoff and the g-band magnitude was greater than 22, then we considered this a true-positive identification. If the "plya_classification" was greater than the cutoff and the g-band magnitude was less than 22, then we considered this a false-positive identification. If the "plya_classification" was less than the cutoff and the g-band magnitude was less than 22, then we considered this a true-negative identification. If the "plya_classification" was less than the cutoff and the g-band magnitude was greater than 22, then we considered this a false-negative identification. The true-positive rate was calculated as true positives divided by the sum of true positives and false negatives. The true-negative rate was calculated as true negatives divided by the sum of true negatives and false positives. The false-positive rate is 1 − true-positive rate, and the false-negative rate is 1 − true-negative rate.

After determining a cutoff for "plya_classification," we then proceeded to examine each source in this redshift range by eye.

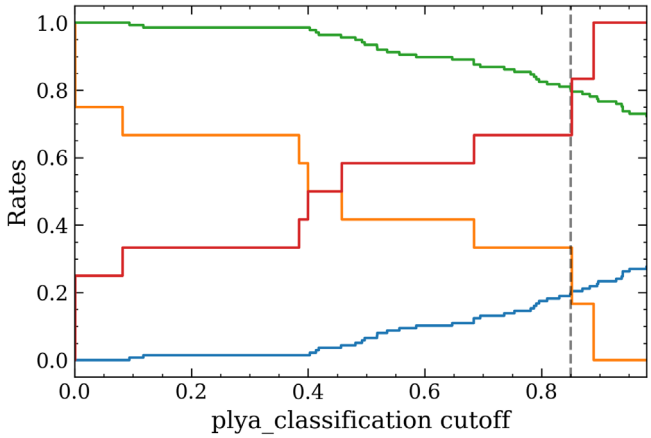

Figure 12. False-positive (blue), false-negative (orange), true-positive (green), and true-negative (red) rates as a function of "plya_classifcation" cutoff. The gray dashed line represents a cutoff of 0.85.

Through this investigation, we determined three patterns to define which redshift to use. Figure 13 shows the three example cases. Using these three patterns, we then determined that, for sources that fall within $2.0 < z_{\rm HDR4} < 3.5$ and $0.0 < z_{\rm Diagnose} < 0.5$, if the "plya_classification" < 0.85, we

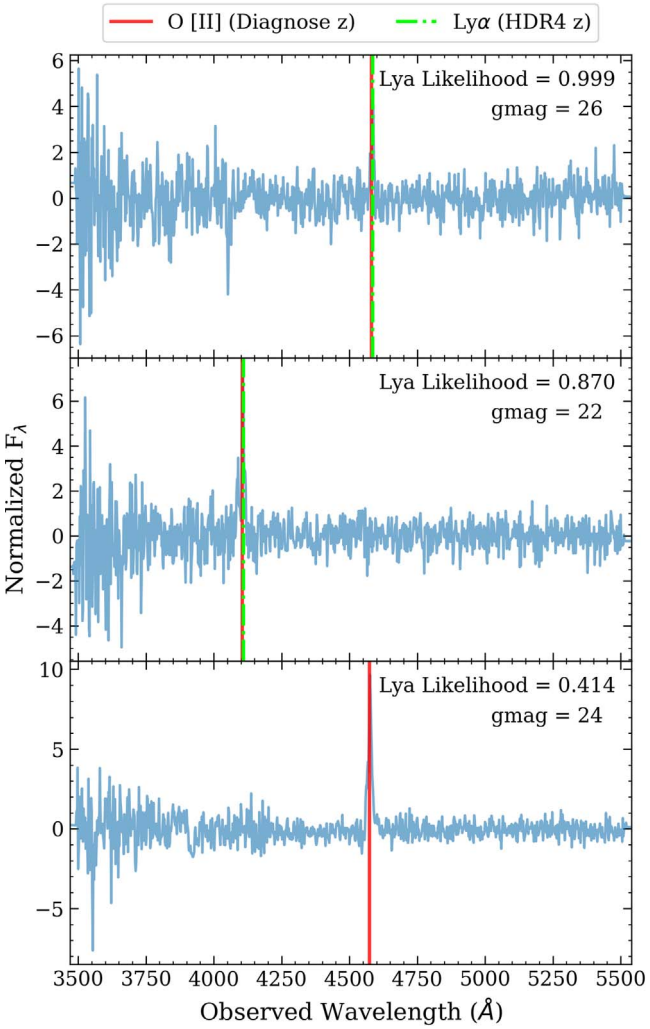

Figure 13. Three cases of redshift determination for $2.0 < z_{\rm HDR4} < 3.5$ and $0.0 < z_{\rm Diagnose} < 0.5$. The top plot shows the case where "plya_classification" > 0.85 and the g-band magnitude is greater than 22, resulting in the use of the HDR4 redshift. The middle plot shows the case where "plya_classification" > 0.85 but the g-band magnitude is bordering 22, resulting in the use of the Diagnose redshift. The bottom plot shows the case where "plya_classification" < 0.85, resulting in the use of the Diagnose redshift.

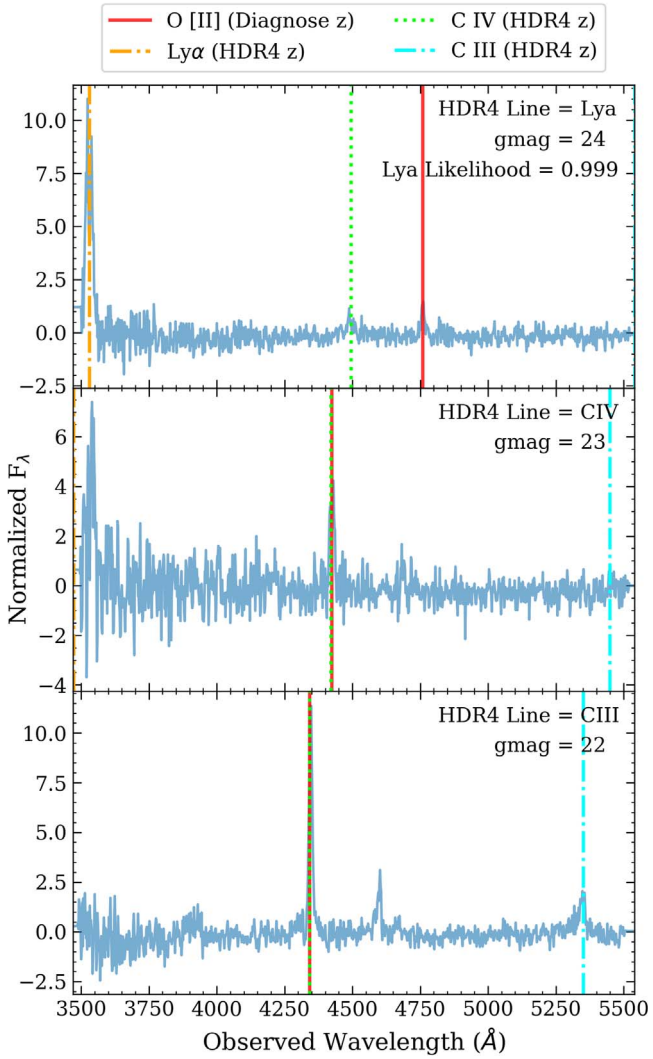

Figure 14. Three cases of redshift determination for $1.5 < z_{\rm HDR4} < 2.0$ and $0.0 < z_{\rm Diagnose} < 0.5$. The top plot shows the case where "plya_classification" > 0.85 and the g-band magnitude is greater than 22, resulting in the use of the HDR4 redshift. The middle plot shows the case where the g-band magnitude is greater than 22, resulting in the use of the HDR4 redshift. The bottom plot shows the case where the g-band magnitude is bordering 22, resulting in the use of the Diagnose redshift.

accept the Diagnose redshift. If "plya_classification" > 0.85 but the g-band magnitude is less than 22, we accept the Diagnose redshift. If "plya_classification" > 0.85 and the g-band magnitude is greater than 22, we accept the HDR4 redshift.

There were 20 sources with $1.5 < z_{\rm HDR4} < 2.0$ and $0.0 < z_{\rm Diagnose} < 0.5$. Some sources in this range also demonstrate the common issue of Diagnose identifying an emission line as [O II] while HDR4 identifies the line as Lyα. In order to distinguish which is the best-fit redshift for this group, we utilized the HDR4 catalog output "plya_classification," which represents the likelihood of the detected line is Lyα when HDR4 identifies the emission line as Lyα. After investigating by eye, we determined three patterns to define which redshift to use. Figure 14 shows the three example cases. Using these three patterns, we then determined that, for sources that fall within $1.5 < z_{\rm HDR4} < 2.0$ and $0.0 < z_{\rm Diagnose} < 0.5$, if the line identified is Lyα and "plya_classification" < 0.85, we accept the Diagnose redshift. If the line identified is Lyα and "plya_classification" > 0.85 but the g-band magnitude is less than 22, we accept the Diagnose redshift. If the line identified is Lyα and "plya_classification" > 0.85 and the g-band magnitude is greater than 22, we accept the HDR4 redshift. If the line identified is not Lyα, then we determine solely based on g-band magnitude. If the g-band magnitude is greater than 22, we accept the HDR4 redshift, and if the g-band magnitude is less than 22, we accept the Diagnose redshift.

There were 18 sources with $1.5 < z_{\rm HDR4} < 2.5$ and $1.0 < z_{\rm Diagnose} < 2.0$. In order to distinguish which is the best-fit redshift for this group, we utilized the HDR4 catalog output "agn_flag," which represents the confidence in the HDR4 AGN classification. A score of 1.0 is a confident AGN, 0.0 is a broadline source but unconfirmed AGN, and -1.0 means it is not an AGN. After investigating by eye, we determined three patterns to define which redshift to use. Figure 15 shows the three example cases. Using these three patterns, we then determined that, for sources that fall within $1.5 < z_{\rm HDR4} < 2.5$ and $1.0 < z_{\rm Diagnose} < 2.0$, if "agn_flag" $= 1.0$, we

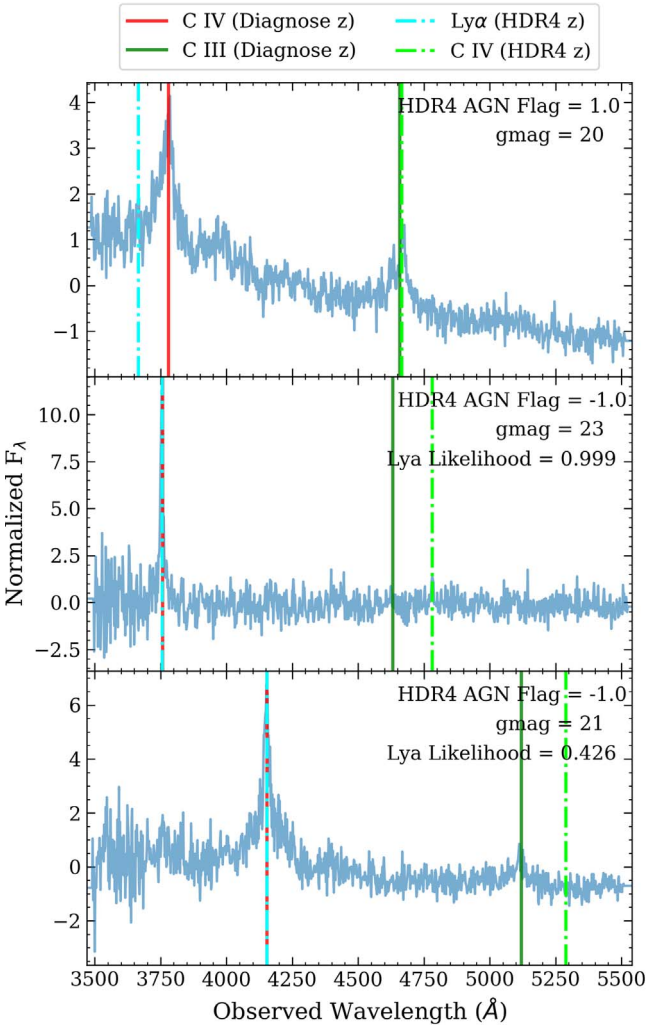

Figure 15. Three cases of redshift determination for $1.5 < z_{\rm HDR4} < 2.5$ and $1.0 < z_{\rm Diagnose} < 2.0$. The top plot shows the case where "agn_flag" = 1.0, resulting in the use of the HDR4 redshift. The middle plot shows the case where "plya_classification" > 0.85 and the *g*-band magnitude is greater than 22, resulting in the use of the HDR4 redshift. The bottom plot shows the case where "plya_classification" < 0.85, resulting in the use of the Diagnose redshift.

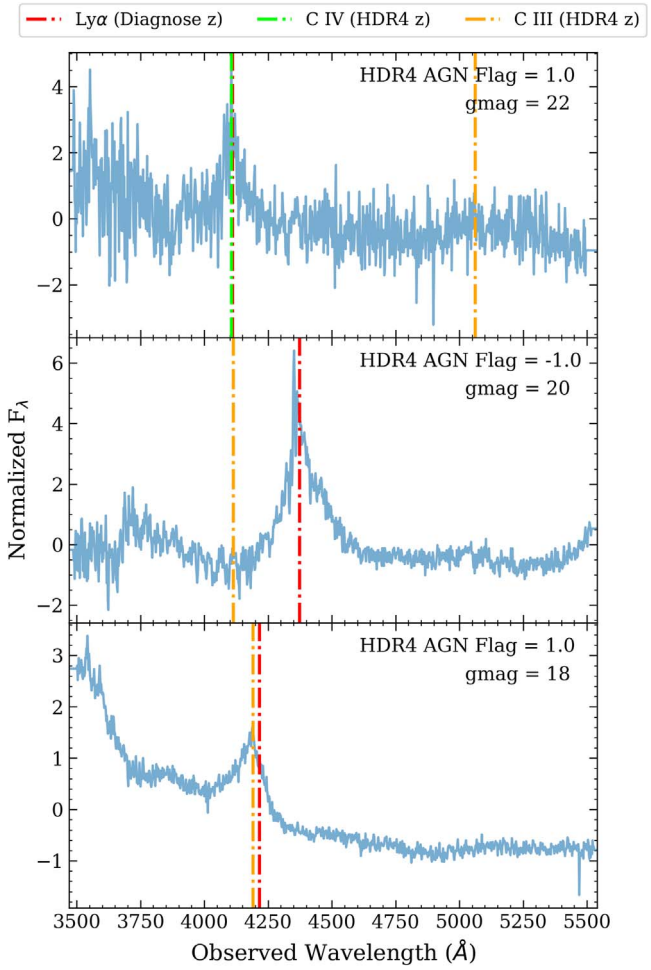

Figure 16. Two cases of redshift determination for $1.0 < z_{\rm HDR4} < 2.0$ and $2.0 < z_{\rm Diagnose} < 3.0$. The top and bottom plots show the case where "agn_flag" = 1.0, resulting in the use of the HDR4 redshift. The middle plot shows the case where "agn_flag" = −1.0 and the *g*-band magnitude is less than 22, resulting in the use of the Diagnose redshift.

accept the HDR4 redshift. If "agn_flag" = −1.0 and "plya_classification" > 0.85 but the *g*-band magnitude is less than 22, we accept the Diagnose redshift. If "agn_flag" = −1.0 and "plya_classification" > 0.85 and the *g*-band magnitude is greater than 22, we accept the HDR4 redshift.

There were 10 sources with $1.0 < z_{\rm HDR4} < 2.0$ and $2.0 < z_{\rm Diagnose} < 3.0$. In order to distinguish which is the best-fit redshift for this group, we utilized the HDR4 catalog output "agn_flag," which represents the confidence in the HDR4 AGN classification. After investigating by eye, we determined two patterns to define which redshift to use. Figure 16 shows the two example cases. Using these two patterns, we then determined that, for sources that fall within $1.0 < z_{\rm HDR4} < 2.0$ and $2.0 < z_{\rm Diagnose} < 3.0$, if "agn_flag" = 1.0, we accept the HDR4 redshift. If "agn_flag" = −1.0 and the *g*-band magnitude is less than 22, we accept the Diagnose redshift. If "agn_flag" = −1.0 and the *g*-band magnitude is greater than 22, we accept the HDR4 redshift.

We investigated an additional category of 25 sources with $0.0 < z_{\rm HDR4} < 0.5$ and $1.0 < z_{\rm Diagnose} < 3.5$. These sources could not follow a specific set of criteria due to the large range of Diagnose redshifts. Figure 17 shows the spectrum for each object in this group.

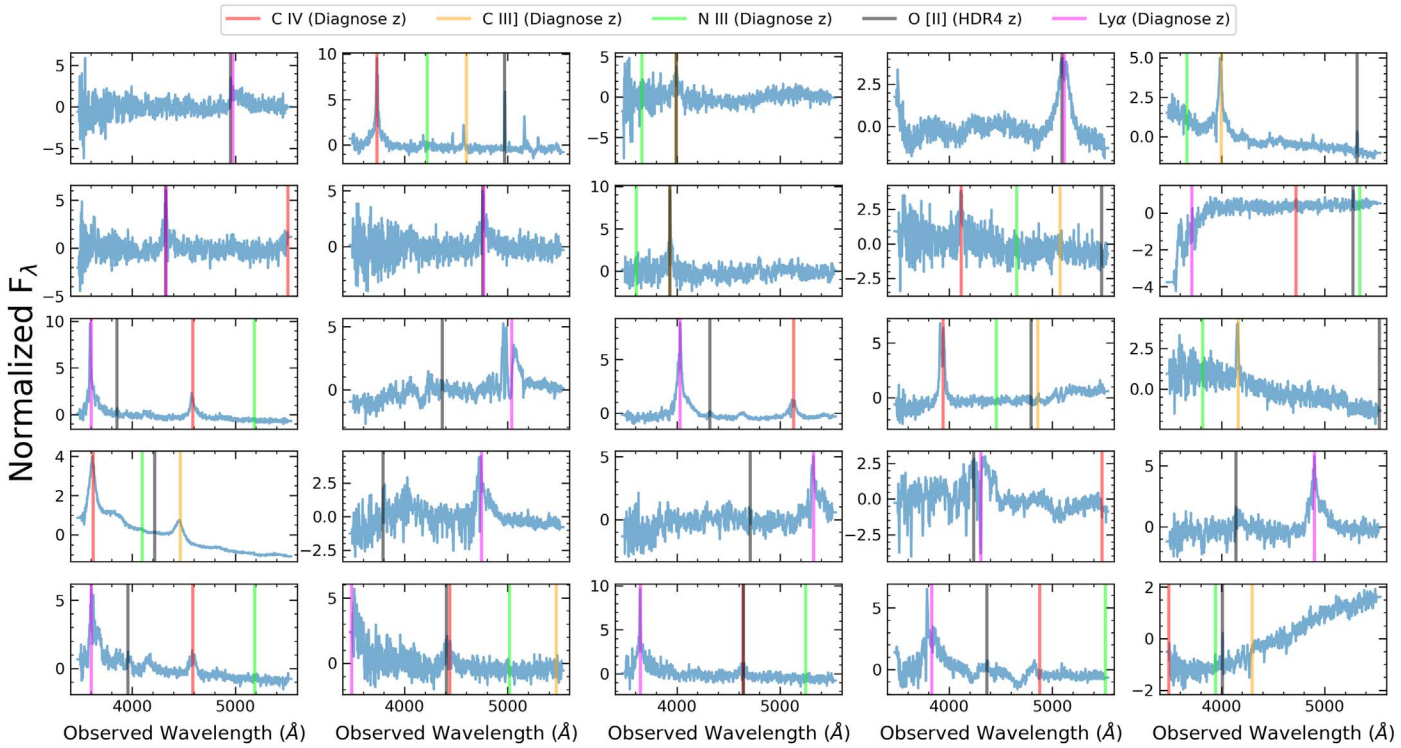

Figure 17. Plots of redshift determination for $0.0 < z_{\rm HDR4} < 0.5$ and $1.0 < z_{\rm Diagnose} < 3.5$. Each colored vertical line represents a different emission-line location determined by either the Diagnose or HDR4 redshift. To classify each of these objects, we examined the line detected by HDR4, which was dominated by [O II] and Lyα. If by optical examination we did not find the HDR4 line at a significant feature, we then examined emission lines from Diagnose. In all of these cases, we found the g-band magnitude to be less than 22 and the Diagnose emission lines to be better representative of the spectra than the HDR4 determined lines.

ORCID iDs

Maya H. Debski https://orcid.org/0000-0002-1998-5677
Gregory R. Zeimann https://orcid.org/0000-0003-2307-0629
Gary J. Hill https://orcid.org/0000-0001-6717-7685
Donald P. Schneider https://orcid.org/0000-0001-7240-7449
Leah Morabito https://orcid.org/0000-0003-0487-6651
Gavin Dalton https://orcid.org/0000-0002-3031-2588
Matt J. Jarvis https://orcid.org/0000-0001-7039-9078
Erin Mentuch Cooper https://orcid.org/0000-0002-2307-0146
Robin Ciardullo https://orcid.org/0000-0002-1328-0211
Eric Gawiser https://orcid.org/0000-0003-1530-8713
Nika Jurlin https://orcid.org/0000-0003-4270-5968

References

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42
Alloin, D., Bergeron, J., & Pelat, D. 1978, A&A, 70, 141
Almeida, A., Anderson, S. F., Argudo-Fernández, M., et al. 2023, ApJS, 267, 44
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, A&AS, 106, 275
Best, P. N., Ker, L. M., Simpson, C., Rigby, E. E., & Sabater, J. 2014, MNRAS, 445, 955
Best, P. N., Kondapally, R., Williams, W. L., et al. 2023, MNRAS, 523, 1729
Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144
Bonzini, M., Padovani, P., Mainieri, V., et al. 2013, MNRAS, 436, 3759
Botteon, A., van Weeren, R. J., Brunetti, G., et al. 2020, MNRAS, 499, L11
Bowman, W. P., Zeimann, G. R., Nagaraj, G., et al. 2020, ApJ, 899, 7
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cappellari, M. 2023, MNRAS, 526, 3273
Chabrier, G. 2003, PASP, 115, 763
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Condon, J. J., Cotton, W. D., & Broderick, J. J. 2002, AJ, 124, 675
Conroy, C., & Gunn, J. E., 2010 FSPS: Flexible Stellar Population Synthesis, Astrophysics Source Code Library
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Cordun, C. M., Timmerman, R., Miley, G. K., et al. 2023, A&A, 676, A29
Dalton, G., Trager, S., Abrams, D. C., et al. 2014, Proc. SPIE, 9147, 91470L
Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, Proc. SPIE, 8446, 84460P
Das, S., Smith, D. J. B., Haskell, P., et al. 2024, MNRAS, 531, 977
Davies, G. R., Lund, M. N., Miglio, A., et al. 2017, A&A, 598, L4
Davis, D., Gebhardt, K., Cooper, E. M., et al. 2023, ApJ, 946, 86
Debski, M., & Zeimann, G. 2024, Diagnose, v1, Zenodo, doi:10.5281/zenodo.13755510
De Zotti, G., Bonato, M., & Cai, Z.-Y. 2019, in Proc. of the Polish Astronomical Soc., Vol. 9, 3rd Cosmology School, Introduction to Cosmology, ed. K. Bajan, M. Biernacka, & A. Pollo, 125
Draine, B. T., & Li, A. 2007, ApJ, 657, 810
Duncan, K. J., Sabater, J., Röttgering, H. J. A., et al. 2019, A&A, 622, A3
Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, A&ARv, 20, 54
Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137
Foreman-Mackey, D., Farr, W., Sinha, M., et al. 2019, JOSS, 4, 1864
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gebhardt, K., Mentuch Cooper, E., Ciardullo, R., et al. 2021, ApJ, 923, 217
Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371
Gloudemans, A. J., Duncan, K. J., Röttgering, H. J. A., et al. 2021, A&A, 656, A137
Grainge, K., Alachkar, B., Amy, S., et al. 2017, ARep, 61, 288
Gürkan, G., Hardcastle, M. J., Smith, D. J. B., et al. 2018, MNRAS, 475, 3010
Heesen, V., Brinks, E., Leroy, A. K., et al. 2014, AJ, 147, 103
Hill, G. J., Lee, H., MacQueen, P. J., et al. 2021, AJ, 162, 298
Horne, K. 1986, PASP, 98, 609
Kukreti, P., Morganti, R., Tadhunter, C., & Santoro, F. 2023, A&A, 674, A198
Levesque, E. M., & Richardson, M. L. A. 2014, ApJ, 780, 100
Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883
Mentuch Cooper, E., Gebhardt, K., Davis, D., et al. 2023, ApJ, 943, 177
Mingo, B., Croston, J. H., Best, P. N., et al. 2022, MNRAS, 511, 3250
Mingo, B., Croston, J. H., Hardcastle, M. J., et al. 2019, MNRAS, 488, 2701
Moffat, A. F. J. 1969, A&A, 3, 455

Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Nagao, T., Maiolino, R., & Marconi, A. 2006, A&A, 459, 85
Padovani, P., Bonzini, M., Kellermann, K. I., et al. 2015, MNRAS, 452, 1263
Padovani, P., Miller, N., Kellermann, K. I., et al. 2011, ApJ, 740, 20
Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, ApJL, 698, L116
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Ramsey, L. W., Adams, M. T., Barnes, T. G., et al. 1998, Proc. SPIE, 3352, 34
Sabater, J., Best, P. N. & LOFAR Collaboration 2019, in Proc. of the XIII Scientific Meeting of the Spanish Astronomical Soc., Highlights on Spanish Astrophysics X, ed. B. Montesinos et al., 231
Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, A&A, 598, A104
Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622, A1
Slee, O. B., Roy, A. L., Murgia, M., Andernach, H., & Ehle, M. 2001, AJ, 122, 1172
Smith, D. J. B., Best, P. N., Duncan, K. J., et al. 2016, in Proc. of the Annual meeting of the French Society of Astronomy and Astrophysics, SF2A-2016, ed. C. Reylé et al., 271
Smith, D. J. B., Haskell, P., Gürkan, G., et al. 2021, A&A, 648, A6
Timmerman, R., van Weeren, R. J., Botteon, A., et al. 2022, A&A, 668, A65
Timmerman, R., van Weeren, R. J., Botteon, A., et al. 2024, A&A, 687, A31,
Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2007, A&A, 461, 823
Whyley, A., Randall, S. W., Clarke, T. E., et al. 2024, arXiv:2402.04876
Williams, W. L., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622, A2
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Wylezalek, D., Galametz, A., Stern, D., et al. 2013, ApJ, 769, 79
Yue, M., Eilers, A.-C., Simcoe, R. A., et al. 2023, ApJ, 950, 105
Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, ApJ, 554, 803
Zeimann, G. R., Debski, M. H., Schneider, D. P., et al. 2024, ApJ, 966, 14