Draft version August 26, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

Identifying Multi-wavelength Counterparts to X-ray Luminous Active Galactic Nuclei in the Hawai'i Twenty Square Degree Survey

Trevin Lee [1,2] Connor Auge,[1] and David Sanders[1]

[1] *Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*
[2] *University of California, Riverside, 900 University Ave, Riverside, CA 92521, USA*

ABSTRACT

Active Galactic Nuclei (AGN) are among the most energetic and influential objects in the universe, driven by supermassive black holes at the centers of galaxies. This study utilizes multi-wavelength data from the Hawai'i Twenty Square Degree Survey (H20), the Chandra Deep Field-South (CDFS), and the SRG/eROSITA all-sky survey to identify and analyze AGN populations across different cosmic epochs. CDFS and eROSITA were matched using a maximum likelihood ratio technique. 149 likely AGN were identified by matching X-ray sources from eROSITA and CDFS with optical and near-infrared sources from H20. The photometric and spectroscopic redshifts from EDFF and CDFS were compared to validate the matching process. The AGN exhibited diverse X-ray luminosities ranging from 42.5 to 45 orders of magnitude in erg/s and a redshift range from 0.23 to 0.44 Higher luminosity AGN displayed more direct UV emissions.

1. INTRODUCTION

Active galactic nuclei (AGN) are among the most luminous and energetic objects in the universe, often outshining their host galaxies. These extraordinary powerhouses are driven by accretion processes occurring around supermassive black holes (SMBH) that reside at the centers of most massive galaxies. As matter is drawn toward the SMBH, it forms an accretion disk where it is heated to extreme temperatures, emitting radiation across the entire electromagnetic spectrum. This emission can have profound effects on the surrounding galaxy, influencing star formation rates, galaxy morphology, and the distribution of gas and dust (see reviews by Heckman & Best (2014); Kormendy & Ho (2013)).

The study of AGN is crucial for understanding a wide range of astrophysical phenomena. AGN are not only key to understanding the growth and evolution of SMBHs, but they also play a significant role in the evolution of their host galaxies. The feedback mechanisms associated with AGN, such as jets and winds, can regulate star formation by heating or expelling gas from galaxies, a process that is critical to models of galaxy evolution.

Over the past few decades, the advent of large-scale astronomical surveys and the development of advanced observational techniques have revolutionized the study of AGN. Surveys across different wavelengths—from radio to X-rays—have allowed astronomers to identify and catalog vast numbers of AGN, providing a comprehensive view of these objects across cosmic time. These surveys have enabled the discovery of a wide variety of AGN types, each offering unique insights into the AGN phenomenon.

One of the major challenges in AGN research is understanding the diversity of AGN properties and their evolution over cosmic time. AGN are observed in a variety of states and environments, ranging from radio-loud quasars in dense galaxy clusters to radio-quiet Seyfert galaxies in more isolated environments (Padovani et al. 2017). This diversity is likely due to differences in SMBH mass, accretion rate, and the properties of the surrounding galactic environment. To unravel these complexities, it is essential to analyze AGN across multiple wavelengths and to combine data from different surveys to build a more complete picture of AGN populations.

This study aims to address these challenges by combining and analyzing data from three large extragalactic catalogs: the Hawai'i Twenty Square Degree Survey (H20), the Chandra Deep Field-South (CDFS), and the SRG/eROSITA all-sky survey. By matching X-ray data from CDFS and eROSITA with optical and near-infrared data from H20, we can perform a detailed statistical analysis of AGN properties, distribution, and evolution across different cosmic epochs. This multi-wavelength approach not only allows us to identify AGN with greater accuracy but also enables us to explore the

connections between AGN activity and the properties of their host galaxies.

The specific objectives of this study include: (1) accurately matching X-ray sources from the eROSITA and CDFS surveys with optical/infrared sources from the H20 survey, (2) validating the reliability of these matches through comparisons of photometric and spectroscopic redshifts, (3) identifying AGN based on their X-ray luminosities, and (4) constructing spectral energy distributions (SEDs) for each identified AGN to analyze their physical properties. By achieving these goals, this study will contribute to a deeper understanding of AGN populations and their role in cosmic evolution.

2. DATA

The datasets utilized in this study are drawn from three major extragalactic surveys: the Hawai'i Twenty Square Degree Survey (H20) (Collaboration et al. 2024), the Chandra Deep Field-South (CDFS) (Luo et al. 2017), and the SRG/eROSITA all-sky survey (Merloni et al. 2024). Together, these datasets provide a multi-wavelength view of the near-infrared, to the X-ray necessary for accurately classifying and analyzing AGN populations. The combination of optical, infrared, ultraviolet, and X-ray data allows for a more complete understanding of the physical properties and evolution of AGN.

2.1. *Hawai'i Twenty Square Degree Survey (H20)*

The H20 Survey (Collaboration et al. 2024) offers extensive coverage over a twenty square degree area, including the North Ecliptic Pole and the Euclid Deep Field Fornax (EDFF). The EDFF catalog provides AB magnitudes and fluxes along with associated errors in the near-infrared to ultraviolet spectrums for over 1,062,645 sources, along with estimated photometric redshifts for objects in a four-square-degree area within the EDFF. The depth and breadth of the H20 catalog make it an ideal survey for studying the large-scale structure and properties of galaxies, particularly when combined with X-ray data to identify AGN.

2.2. *SRG/eROSITA All-Sky Survey*

The eROSITA survey (Merloni et al. 2024) is the most recent and comprehensive X-ray survey, covering a majority of the sky. eROSITA contains 930,203 X-ray sources, with 221 sources overlapping with the H20 field. eROSITA predominantly provides X-ray fluxes in the 0.5 keV to 2.3 keV range in the 1B catalogue with a flux limit of $F_{0.5-2\,\mathrm{keV}} > 5 \times 10^{-14}$ erg s^{-1} cm^{-2}. This catalog is crucial for identifying AGN based on their X-ray emissions, despite the absence of redshift information as it provides recent data that hasn't been studied as extensively as CDFS has.

2.3. *Chandra Deep Field-South (CDFS)*

The CDFS catalog (Luo et al. 2017) is one of the deepest X-ray surveys ever conducted, with a 7 Ms exposure over a 0.11 square degree field, conveniently located entirely within the area covered by EDFF. The CDFS catalog contains 1,055 sources, with deeper coverage extending further into the night sky. Each source is provided with spectroscopic redshift data and full-band X-ray fluxes from 0.5 keV to 10.0 keV, with flux limits reaching as low as $\sim 7 \times 10^{-18}$ erg s^{-1} cm^{-2} in the soft band (0.5–2 keV) and $\sim 2 \times 10^{-17}$ erg s^{-1} cm^{-2} in the hard band (2–10 keV). This catalog is invaluable for validating the methods used to match X-ray to H20 sources as it provides robust spectroscopic redshifts to compare to the photometric redshifts provided by H20.

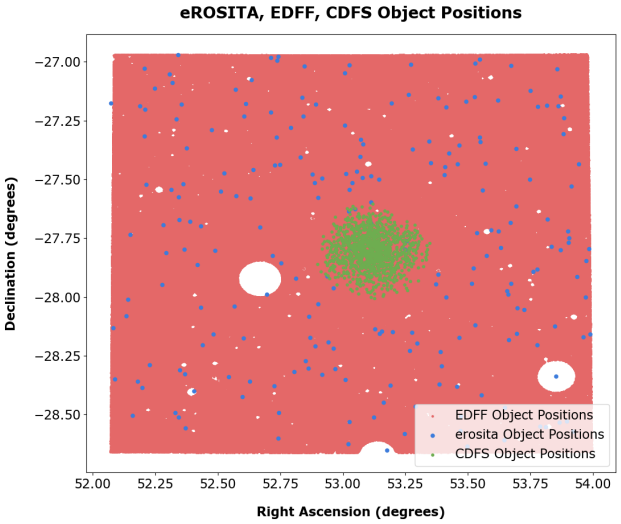

Figure 1. The spatial distribution of sources from the H20, CDFS, and eROSITA catalogs within the EDFF region. This figure illustrates the overlap and coverage of each dataset, highlighting the complementary nature of these surveys for AGN identification.

3. METHODS

This section details the methods used to match and validate the multi-wavelength catalogs, identify AGN, and construct spectral energy distributions (SEDs) for the AGN identified in this study.

3.1. *Catalog Matching*

To classify AGN and perform a detailed statistical analysis, we matched the H20 multi-wavelength catalog

with the X-ray catalogs from eROSITA and CDFS. The matching process was conducted using the LYR script (Peca et al. 2021; Nanni et al. 2020), which is well-suited for handling datasets with varying resolutions and noise levels across different wavelengths. The LYR script employs a maximum likelihood ratio (LR) method to determine the probability that an object in one catalog is the true counterpart of an object in another catalog. The likelihood ratio for a given match is calculated using the formula:

$$\mathrm{LR} = \frac{F(r)q(f)}{n(f)}, \tag{1}$$

where $F(r)$ is the Gaussian probability distribution function of the angular separation between the two objects, and $q(f)$ represents the expected flux distribution of true counterparts. The background object density $n(f)$ is estimated by counting the number of objects within an annulus around the candidate match and normalizing by the area of the annulus. The expected flux distribution $q(f)$ cannot be directly measured but is estimated as:

$$q(f) = \frac{\mathrm{real}(f)}{\sum_{i=1}^{n} \mathrm{real}(f)_i} Q, \tag{2}$$

where $\mathrm{real}(f)$ is the total number of potential counterparts minus a correction for background sources, and Q is the probability that a true counterpart is brighter than the flux limit. The reliability R of a match is then calculated as:

$$R = \frac{\mathrm{LR}}{\sum_i \mathrm{LR}_i + (1 - Q)}, \tag{3}$$

where the sum is over all potential counterparts for a given source, and Q accounts for the flux limit of the catalog. This method ensures a robust identification of counterparts across the different catalogs. Finally the completeness of the catalog can be computed by:

$$C = \frac{\left(\sum_j R_j\right)_{LR > LR_{th}}}{N}. \tag{4}$$

Where N is the total number of objects LYR is attempting to find a match for and $(\sum_i R_i)_{LR>LR_{th}}$ is sum of the reliabilities of the objects for which there is a match. An object is considered to be matched if its likelihood ratio is greater than the likelihood ratio threshold (LR_{th}) which is the sum of the reliability and completeness curves divided by 2. The likelihood ratio threshold for the catalog is set where this value is maximized.

3.2. *Matched Catalog*

For the CDFS catalog, LYR was able to find 2791 matches and 421 matches for eROSITA. The CDFS and eROSITA matches had an LR_{th} of 0.06 and 1.51 respectively. If a match were less than the LR_{th} or had an LR_{th} of less than 1 it was removed from the catalog. All matches with missing data were also removed.

The LYR script yielded a total of 656 matches between the CDFS and EDFF catalogs, of which only 468 contained spectroscopic redshift and soft-band X-ray data, and 53 matches for the eROSITA of which 50 sources contained valid data. If the same EDFF source was matched to both an eROSITA and CDFS source, the CDFS match was taken over the eROSITA match due to the more complete data. This resulted in a total of 513 matches, with 5 duplicate eROSITA matches removed.

3.3. *Catalog Validation*

To validate the accuracy of the catalog matching, we used the CDFS catalog, which provides spectroscopic redshift information not available in the eROSITA catalog. The H20 catalog was matched with the CDFS X-ray catalog, and the redshifts of the matched sources were compared. This validation step involved assessing the precision of the matches by comparing photometric redshifts from the H20 catalog with the more accurate spectroscopic redshifts from the CDFS catalog. Matching in three dimensions—Right Ascension (RA), Declination (DEC), and redshift—provides a more robust verification of the accuracy of our matching process. In addition, the correlation coefficients were computed as well as the normalized median absolute distribution (NMAD) (Hoaglin et al. 1983). The NMAD was compared to the verifications done by (Collaboration et al. 2024) in the original H20 paper.

The duplicate sources shown in 2 also reveal insights into the quality of the matched catalog. Of the 5 duplicate sources, 4 of the 5 sources are nearly identical in their X-ray luminosities as observed by the Chandra and Spectrum-Roentgen-Gamma (SRG) telescopes. Although the sample is quite limited, This further corroborates that the matching performed by LYR was quite accurate and also that the telescopes themselves agree quite well on the measured luminosities.

?? shows the separation between the CDFS and EDFF sources matched by LYR. The mean and standard deviation of the separations are 0.43 and 0.19 arc seconds respectively with the maximum separation at 0.87 arc seconds. The maximum separation of the dataset falls well below the maximum positional error of the sources provided in the CDFS catalog of 1.78 arc seconds. The mean positional error of the dataset was 0.43 and which falls below the mean positional error of 0.49 in the CDFS catalog. Positional errors were not provided in the EDFF catalog, but were assumed to be very

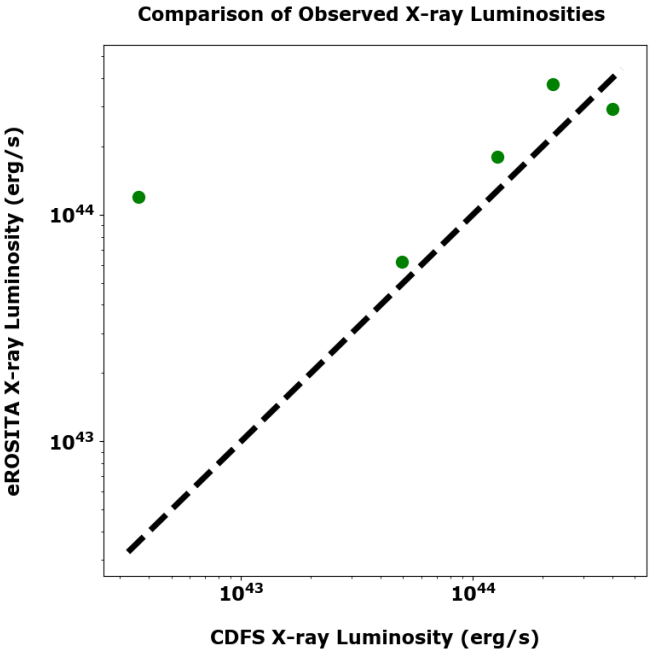

Figure 2. The X-Ray luminosities of the double matched sources as observed in EDFF and CDFS

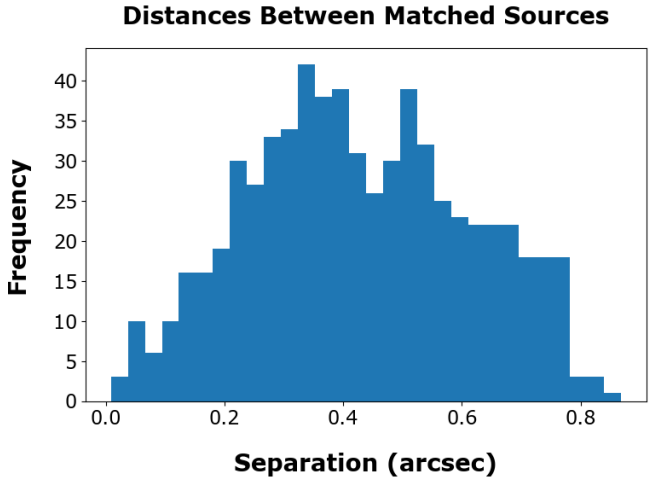

Figure 3. Distribution of Distances Between Matched Sources in EDFF and CDFS

small due to the nature of measurements at higher wavelengths. Finally, 4 shows the photometric and spectroscopic redshifts of the EDFF and CDFS catalogs. It is evident in the plot that a majority of the sources fall very close to the $x = y$ line with many of these sources having high likelihood and reliabilities. The correlation coefficient of the matches were computed to be 0.86 showing a strong linear correlation.

This is strong evidence that LYR is a highly reliable method for matching catalogs as LYR was never provided with redshift data. There exists only one false positive near the bottom right corner of the plot, and still, this may be attributed to an error in the photometric redshift of the source as it exits far below the $x = y$ line.

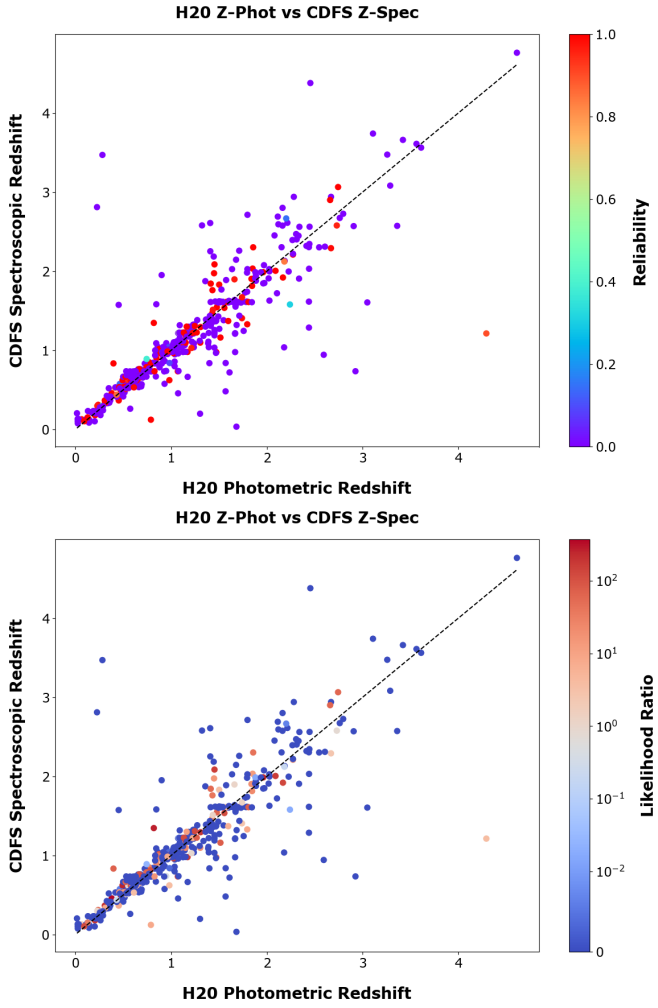

Figure 4. H20 Photometric Redshift vs CDFS Spectroscopic Redshift colored by likelihood ratio and reliability

Furthermore, the matching results are likely more robust than they appear, as photometric redshift estimations include significantly higher error than spectroscopic redshift estimations. Therefore, many of the misaligned data points may be attributed to errors in photometric redshift. In the original H20 paper (Collaboration et al. 2024) the photometric redshifts calculations were verified against provided the spectroscopic redshifts of some

The normalized median absolute distribution (NMAD) of the matched dataset was computed and compared to that of H20 catalog in order to further validate the results. The NMAD Hoaglin et al. (1983)

of the redshifts were computed by:

$$\sigma_{\rm NMAD} = 1.48 \times \text{median}\left(\left|\frac{\Delta z - \text{median}(\Delta z)}{1 + z_{\rm spec}}\right|\right) \quad (5)$$

When comparing the NMAD of the newly matched catalog with that of the original validation done in Collaboration et al. (2024) for the EDFF dataset, the newly matched catalog performs similarly to the NMAD of 0.035 of the EDFF catalog with an NMAD of 0.0925. Although, the NMAD of the newly matched catalog is 2.6 times higher, it is very small at 0.0925 and close enough to the 0.035 NMAD of the EDFF to be considered similar and ultimately to conclude that the many of the differences in photometric and spectroscopic redshifts are due to the errors in photometric redshifts of the EDFF catalog.

3.4. *AGN Identification*

Following the validation of our catalog matches, the two sets of matches of the eROSITA and EDFF catalogs were merged, and AGN were identified primarily using X-ray luminosity as the distinguishing criterion. The X-ray luminosities were calculated using the Flat Lambda Cold Dark Matter (FlatLambdaCDM) cosmological model and spectroscopic redshifts from CDFS and photometric redshifts from H20. All objects with an X-ray luminosity greater than $10^{42.5}$ erg s^{-1} were identified as very likely to be AGN. This threshold is consistent with established literature (Ranalli et al. 2003), and is a conservative method which allowed near certainty that a large majority of the AGN dataset were in fact AGN.

After applying the classification threshold, the dataset contained 102 sources from CDFS and 47 sources from eROSITA, for a total of 149 sources that were very likely to be AGN 5.

3.5. *Spectral Energy Distribution (SED) Construction*

To further analyze the identified AGN, spectral energy distributions (SEDs) were computed for each AGN. SEDs are essential for understanding the physical properties of AGN, including their energy output across different wavelengths. For each AGN, we used the median wavelength value of each telescope filter to represent the flux at a specific wavelength. For the X-ray data, we calculated a median wavelength corresponding to the soft-band (0.5 - 2.0 keV) photon energies. These wavelength values were then corrected to their rest-frame values using the redshifts provided by the CDFS and EDFF catalogs. To facilitate comparison across the sample, all SEDs were interpolated and normalized at a wavelength of 1 micron. This normalization allows for the identification of trends and the comparison of AGN with different intrinsic luminosities and redshifts.

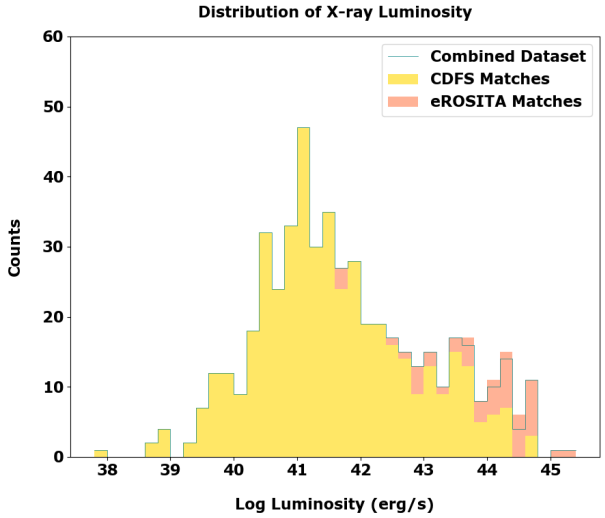

Figure 5. The Distribution of Soft Band X-ray Luminosity of all sources in final merged catalog

4. RESULTS

4.1. *Spectral Energy Distributions (SEDs)*

AGN with lower UV luminosities exhibited SEDs that clustered together, showing a characteristic increase in the optical region followed by a decline towards the infrared. In contrast, AGN with higher X-ray luminosities displayed significantly higher UV emissions and a more scattered distribution in their SEDs 6. 7 shows the SEDs binned by X-ray luminosity and redshift. The pattern

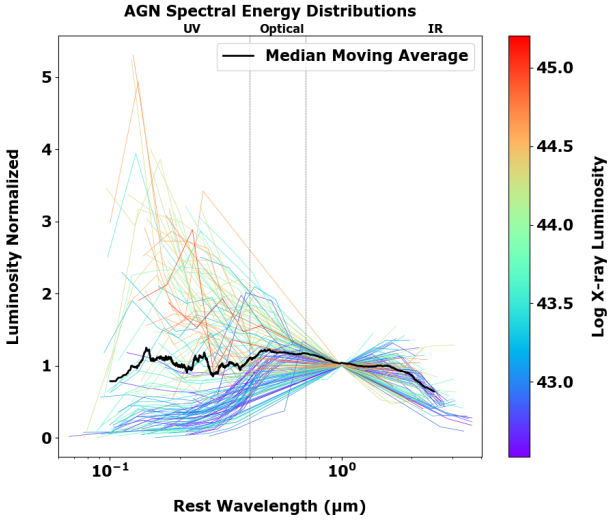

Figure 6. AGN Spectral Energy Distributions Normalized at 1 Micron and colored by X-ray Luminosity

mentioned above is more clearly visible in the 42-43 log luminosity bin. These SEDs are lower X-ray luminosities are clustered together and follow a similar trend with lower UV emissions and slope similar to higher optical emissions with few outliers. In contrast, AGN in the ¿44 bin show very chaotic distributions in the UV and show significantly higher UV emissions which slope downwards towards the optical. The 43-44 log luminosity bin show a combination of SEDs with both characteristics. When analyzing the AGN SEDs sorted by redshifts it seems that the bin contain AGN at redshifts 0-1 are less luminous in the UV.

5. DISCUSSION

5.1. *Spectral Energy Distributions (SEDs)*

The construction of spectral energy distributions (SEDs) for each AGN revealed distinct trends based on X-ray luminosity and redshift. AGN with lower UV luminosities exhibited SEDs that clustered together, with a characteristic rise in the optical region followed by a decline towards the infrared. This pattern is consistent with the presence of significant dust obscuration, which absorbs UV and optical light and re-emits it in the infrared. In contrast, AGN with higher X-ray luminosities displayed significantly higher UV emissions and a more scattered distribution in their SEDs. This suggests that these AGN have less surrounding dust or that their intense radiation fields are capable of clearing out the dust, leading to more direct UV emission. The diversity observed in the SEDs across different AGN indicates a range of physical conditions, such as varying levels of dust content, accretion rates, and SMBH masses. These findings are in line with previous studies that have noted the wide variety of SED shapes among AGN (Hickox & Alexander 2018). The observed trends provide further evidence that AGN play a complex role in galaxy evolution, influencing and being influenced by their environments in diverse ways.

5.2. *Limitations and Future Work*

While this study has provided important insights into AGN populations, there are limitations that should be acknowledged. One minor limitation is the reliance on photometric redshifts from the H20 survey, which, while effective for large-scale studies, are less accurate than spectroscopic redshifts. This limitation was partially mitigated by using the CDFS spectroscopic data for validation, but future studies could benefit from acquiring more spectroscopic redshifts to improve the precision of AGN classifications. Additionally, the study's sample is limited to the data available in the H20, eROSITA, and CDFS surveys. Expanding this work to include data from other regions for greater statistical significance, additional wavelength ranges such as radio or submillimeter observations, and data from far higher redshifts could provide a more comprehensive understanding of AGN across different environments and cosmic epochs.

6. CONCLUSIONS

This study examined Active Galactic Nuclei (AGN) using data from the Hawai'i Twenty Square Degree Survey (H20), the Chandra Deep Field-South (CDFS), and the SRG/eROSITA all-sky survey. By matching and validating catalogs across different wavelengths using LYR, 513 total matches were found and 149 likely AGN were identified, providing a solid foundation for further analysis of their properties.

The study validated the effectiveness of the LYR script applied to the H20 catalogs in accurately matching counterparts between catalogs, which was confirmed by a strong correlation between photometric and spectroscopic redshifts. This validation was essential in ensuring the reliability of the AGN identification process.

Analysis of the identified AGN revealed a diverse population with varying X-ray luminosities and spectral energy distributions (SEDs). AGN with lower X-ray luminosities tended to exhibit clustered SEDs with significant dust obscuration, while those with higher luminosities showed more direct UV emissions. These findings suggest differences in dust content, accretion activity, and possibly the evolutionary stages of the AGN.

The results have important implications for understanding the role of AGN in galaxy evolution. The diversity in AGN luminosities and SEDs indicates that AGN can have multiple influences on star formation and galaxy morphology, depending on their activity levels and environmental conditions. This study contributes to the ongoing research on how AGN impact their host galaxies and the broader universe.

Future research could benefit from additional data at higher redshifts, with higher granularity SEDs, and a larger range of wavelengths could give greater insight into the AGN evolution across cosmic epochs. More provided spectroscopic redshifts would also greatly enhance the accuracy of the catalog matching.

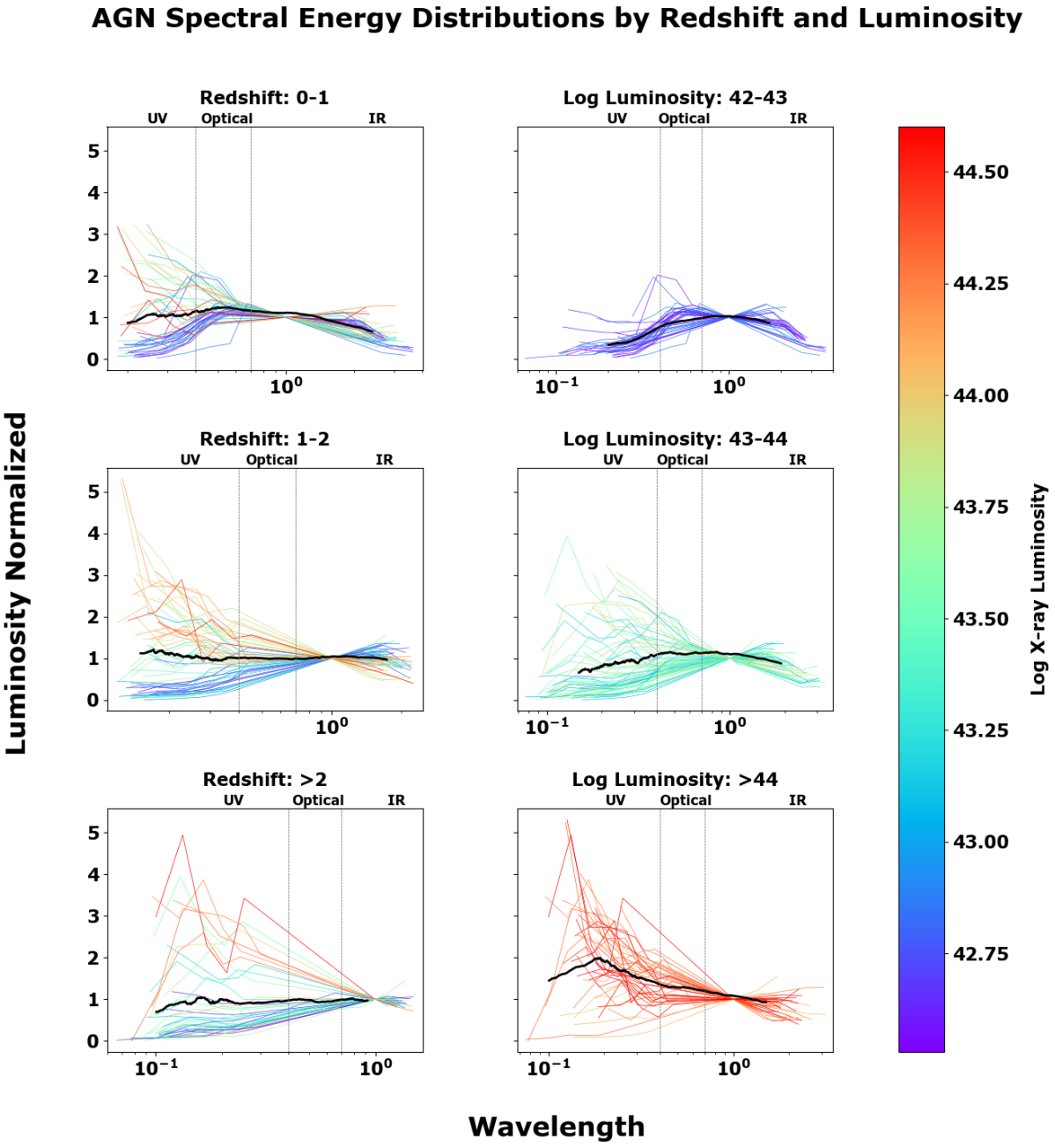

Figure 7. Normalized AGN Spectral Energy Distributions sorted into Luminosity and Redshift Bins

REFERENCES

Collaboration, E., Zalesky, L., McPartland, C. J. R., et al. 2024, Euclid Preparation. Cosmic Dawn Survey: Data release 1 multiwavelength catalogues for Euclid Deep Field North and Euclid Deep Field Fornax. https://arxiv.org/abs/2408.05296

Heckman, T. M., & Best, P. N. 2014, ARA&A, 52, 589, doi: 10.1146/annurev-astro-081913-035722

Hickox, R. C., & Alexander, D. M. 2018, ARA&A, 56, 625, doi: 10.1146/annurev-astro-081817-051803

Hoaglin, D., Mosteller, F., & Tukey, J. 1983, Understanding Robust and Exploratory Data Analysis, WILEY SERIES in PROBABILITY and STATISTICS: PROBABILITY and STATISTICS SECTION Series (Wiley). https://books.google.com/books?id=HRrvAAAAMAAJ

Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511, doi: 10.1146/annurev-astro-082708-101811

Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, ApJS, 228, 2, doi: 10.3847/1538-4365/228/1/2

Merloni, A., Lamer, G., Liu, T., et al. 2024, A&A, 682, A34, doi: 10.1051/0004-6361/202347165

Nanni, R., Gilli, R., Vignali, C., et al. 2020, A&A, 637, A52, doi: 10.1051/0004-6361/202037914

Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&A Rv, 25, 2, doi: 10.1007/s00159-017-0102-9

Peca, A., Vignali, C., Gilli, R., et al. 2021, ApJ, 906, 90, doi: 10.3847/1538-4357/abc9c7

Ranalli, P., Comastri, A., & Setti, G. 2003, A&A, 399, 39, doi: 10.1051/0004-6361:20021600