








The Chandra Source Catalog Normal Galaxy Sample

Dong-Woo Kim¹ , Alyssa Cassity², Binod Bhatt³, Giuseppina Fabbiano¹ , Juan Rafael Martinez Galarza¹ ,
Ewan O’Sullivan¹ , and Arnold Rots¹ 

¹ Center for Astrophysics | Harvard and Smithsonian, MA, USA; dkim@cfa.harvard.edu

² University of British Columbia, Canada

³ University of Rochester, NY, USA

Received 2023 April 6; revised 2023 June 14; accepted 2023 July 3; published 2023 August 25

Abstract

We present the extensive and well-characterized Chandra X-ray Galaxy Catalog (CGC) of 8547 galaxy candidates in the redshift range $z \sim 0.04\text{--}0.7$, optical luminosity $10^{10}\text{--}10^{11} L_{r\odot}$, and X-ray luminosity (0.5–7 keV) $L_X = 2 \times 10^{40}\text{--}2 \times 10^{43} \text{ erg s}^{-1}$. We estimate a $\sim 5\%$ false-match fraction and contamination by quasi-stellar objects (QSOs). The CGC was extracted from the Chandra Source Catalog version 2 (CSC2) by cross-correlating with optical and IR all-sky survey data, including the Sloan Digital Sky Survey, Pan-STARRS, DESI Legacy, and Wide-field Infrared Survey Explorer (WISE). Our selection makes use of two main criteria that we have tested on the subsample with optical spectroscopic identification. (1) A joint selection based on X-ray luminosity (L_X) and X-ray-to-optical flux ratio (F_{XO}), which recovers 63% of the spectroscopically classified galaxies with a small contamination fraction (7%), which is a significant improvement over methods using L_X or F_{XO} alone ($<50\%$ recovery). (2) A joint W1–W2 (W_{12}) WISE color and L_X selection that proves effective in excluding QSOs and improves our selection by recovering 72% of the spectroscopically classified galaxies and reducing the contamination fraction (4%). Of the CGC, 24% was selected by means of optical spectroscopy; 30% on the basis of L_X , F_{XO} , and W_{12} ; and 46% by using either the $L_X\text{--}F_{XO}$ or the $L_X\text{--}W_{12}$ selection criteria. We have individually examined the data for galaxies with $z < 0.1$, which may include more than one CSC2 X-ray source, leading to the exclusion of 110 local galaxies. Our catalog also includes near-IR and UV data and galaxy morphological types.

Unified Astronomy Thesaurus concepts: [Galaxies \(573\)](#)

Supporting material: tar.gz file

1. Introduction

Over the past 20 yr, Chandra X-ray observations have led to a new understanding in several areas of astronomy and astrophysics (see Wilkes & Tucker 2019). The unprecedented high spatial resolution of Chandra⁴ provides the ability to resolve fine structures at a subarcsecond scale and detect the faintest X-ray sources. These capabilities have taken to a new level the study of galaxies and their components: the hot interstellar medium (ISM), populations of low and high-mass X-ray binaries (XRBs), and nuclear activity. These Chandra results have highlighted how the X-ray emission is important for our understanding of the formation and evolution of galaxies (e.g., Kim & Pellegrini 2012; Fabbiano 2019). For example, the hot ISM, which was previously considered as smooth and featureless, has been shown to have complex X-ray morphologies and spectral properties, pointing to fundamental evolutionary mechanisms, such as active galactic nuclei (AGNs) feedback, merging history, accretion, stripping, and star formation (SF) and its quenching (Fabbiano & Elvis 2022; Nardini et al. 2022). Similarly, the evolution of XRBs, as well as AGNs, may also result in substantial energy input (feedback) into the ISM (Fabian 2012; Choi et al. 2017; Ciotti et al. 2017).

The aim of this work is to build an extensive, unbiased catalog of X-ray-selected normal (not AGN-dominated) galaxies, the Chandra X-ray Galaxy Catalog (CGC). The CGC is needed to place the scaling relations between observed parameters (e.g., L_X , T , multiwavelength luminosity, and colors) and fundamental physical parameters of galaxies (e.g., mass and SF) on a stronger statistical footing. These parameters have been discussed in the literature (Boroson et al. 2011; Li & Wang 2013, Kim & Fabbiano 2015; Bogdan & Goulding 2015). The catalog is also needed to produce the highly significant X-ray luminosity function of galaxies. While the Chandra observations of nearby galaxies have been instrumental in producing the luminosity function of XRB populations and thereby exploring the evolution of these binary stellar systems (e.g., Kim et al. 2009; Kim & Fabbiano 2010; Mineo et al. 2012; Lehmer et al. 2016), it has not been possible to fully explore the X-ray luminosities of different types of galaxies because the available sample size was limited (e.g., Kim et al. 2006). Numerical simulations can broadly reproduce the scaling relations of hot-gas properties (e.g., $L_{X,GAS}$ and T_{GAS}), emphasizing the importance of stellar and AGN feedback in galaxy formation and evolution (Choi et al. 2017; Ciotti et al. 2017; Vogelsberger et al. 2020; Kelly et al. 2021). The CGC provides the means to place significantly stricter constraints on the observed parameter space.

The Chandra Multiwavelength Project (ChaMP) conducted the first serendipitous survey of the sky with Chandra, producing ~ 7000 sources from a sky area of 10 deg^2 (Kim et al. 2007), which were identified with stars, galaxies, AGNs, and galaxy clusters. A minority of these sources, 136,

⁴ <https://asc.harvard.edu/proposer/POG/>

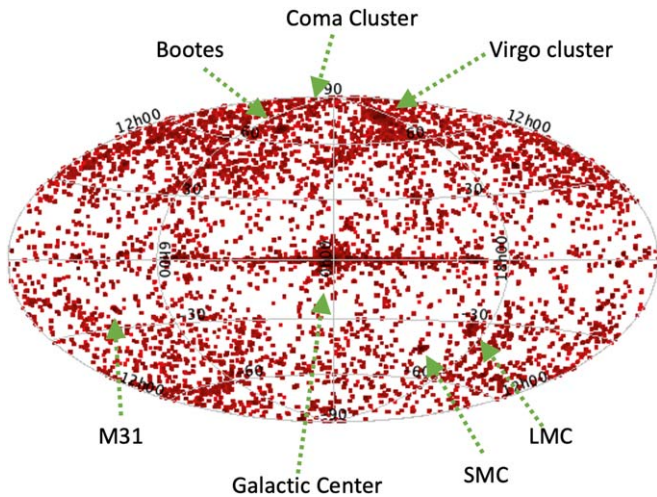


Figure 1. X-ray sources in CSC2 are plotted in Galactic coordinates (l , b). A few well-known objects are also marked.

were identified with normal galaxies (Kim et al. 2006), a result consistent with that of the COSMOS X-ray survey, where normal galaxies comprise about a few percent of the sample (Marchesi et al. 2016). Therefore, to increase the normal galaxy sample significantly, we need to survey a much larger region of the sky. The Chandra Source Catalog Version 2 (CSC2) provides this opportunity. The CSC2 includes 317,167 unique X-ray sources with accurate positions and high-quality photometric, spectral, and temporal properties (Evans et al. 2010). To extract an expanded sample of normal galaxies, we have crossmatched the CSC2 with a suite of optical, IR, and UV catalogs. Our selection makes use of accurate positions and selection criteria aimed at minimizing sample contamination by AGNs and QSOs.

This paper is organized as follows. In Section 2, we describe the X-ray source selection. In Section 3 we describe the optical, IR, and UV catalogs that we used to crossmatch X-ray sources and classify galaxies. In Section 4 we present our crossmatching procedures and the match statistics, including false-match rates. In Section 5 we describe the galaxy-finding algorithms and the galaxy properties in the multiwavelength parameter space. In Section 6 we present the final CGC. In Section 7 we list the science topics that we plan to explore with our CGC.

Throughout the paper, we adopt the following cosmological parameters: $H_0 = 69.6 \text{ (km s}^{-1}\text{) Mpc}^{-1}$, $\Omega_M = 0.286$, and $\Omega_\Lambda = 0.714$.

2. The Chandra X-Ray Source Catalog

The main advantage of the Chandra Observatory is its unprecedented high spatial resolution—the capability to resolve individual sources in a crowded field and to determine their positions on a subarcsecond scale. This resolution also results in the best detection efficiency for the faintest X-ray sources in the sky. The CSC2,⁵ released in 2019, makes use of the first 15 yr of Chandra observations (1999–2014), coadded in overlapping regions of the sky to achieve the best sensitivity. The CSC2 includes 317,167 unique X-ray sources with accurate positions and high-quality photometric, spectral, and temporal properties. Because of the angular resolution of Chandra, the CSC2 provides the most precise source position and is least affected

by source confusion among the X-ray catalogs. Figure 1 shows the CSC2 sources plotted in Galactic coordinates in the Aitoff projection.

Out of the 317,167 CSC2 sources, we selected only those with high data quality by applying the following additional filters:

1. Streak flag = False (to exclude sources located on an ACIS readout streak).
2. Pile-up flag = False (to exclude sources with an ACIS pile-up fraction >10%).
3. Saturation source flag = False (to exclude sources for which all stacked observation detections are significantly piled up).

Moreover, to minimize the false-association rate to $\sim 5\%$, we excluded crowded sky regions ($|b| < 15^\circ$, and higher Galactic latitude crowded fields, typically including Local Group galaxies or Galactic star clusters—see Section 4). We excluded 1299 highly extended ($>30''$) sources, called convex-hull sources.⁶ Their source properties are considered as an alpha-release in CSC2 and are not as reliable as the compact sources. We also excluded the CSC2 sources for which the positions of the optical or IR counterparts differed by more than 1 or 2'' (see Section 4 for details). This screening resulted in a catalog of 117,358 high-quality high-Galactic-latitude pointlike or moderately extended sources.

3. Multiwavelength Catalogs

We crossmatched the X-ray sources with optical, IR, and UV catalogs to identify galaxies and obtain multiwavelength data using the all-sky surveys listed below. To illustrate the sky coverage of each catalog, we show the matched sources from each catalog in Figure 2 in Galactic coordinates in the Aitoff projection.

3.1. SDSS

The⁷ spectroscopic and photometric catalogs in DR16 include 5 million and 800 million sources in the SpecObj and PhotoObj tables, respectively. In the spectroscopic sample, $\sim 900,000$ galaxies have additional information on their morphological types from the Galaxy Zoo.⁸ Those in the PhotoObj table (measured in five bands— u , g , r , i , and z) consist of 430 million pointlike objects (type = 6 or star) and 370 million extended objects (type = 3 or galaxy). In the photometric sample, the photometric redshifts of ~ 200 million objects are available (Beck et al. 2016)

3.2. Pan-STARRS

The DR2⁹ includes 3 billion (bn) sources from the 3π sky with $\delta > -30^\circ$, with optical photometric data in five bands (g , r , i , z , y) and surface brightness fits to morphological models (e.g., exponential and de Vaucouleurs). Although their spectroscopic redshifts are not available, photometric redshifts are available for 2.9 bn sources (Beck et al. 2021). Additionally, 1.6 million galaxies have broad morphological types by Goddard & Shamir (2020).

⁵ <https://cxc.harvard.edu/csc/>

⁶ https://cxc.harvard.edu/csc/columns/chs_properties.html

⁷ <https://www.sdss.org/dr16/>

⁸ <https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/>

⁹ <https://panstarrs.stsci.edu>

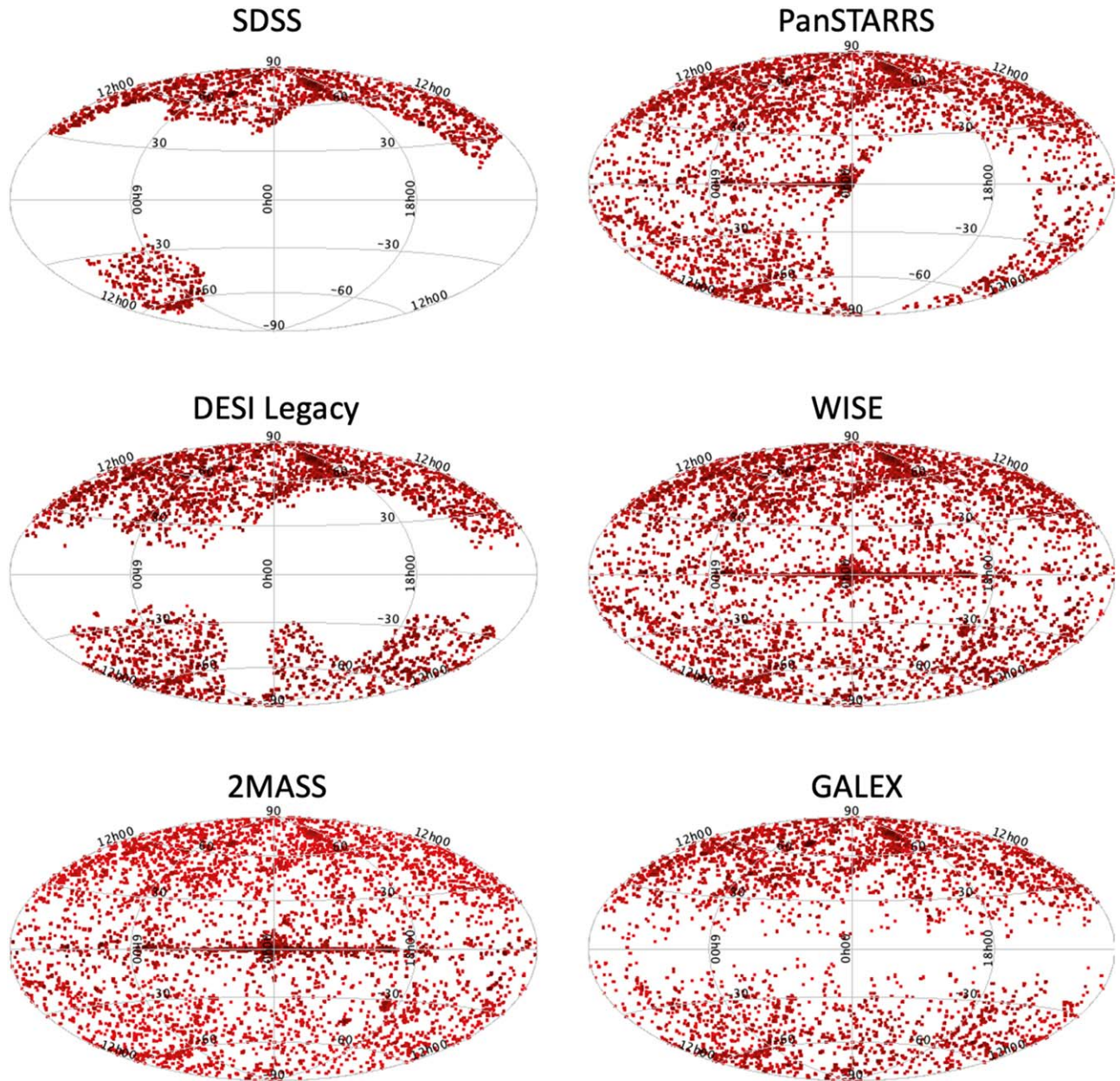


Figure 2. The optical, IR, and UV counterparts of CSC2 sources plotted in Galactic coordinates (l, b).

3.3. DESI Legacy

The DR8¹⁰ catalog includes 2 bn sources from $14,000 \text{ deg}^2$ of the extragalactic sky in $-18^\circ < \delta < +84^\circ$ and $|b| > 18^\circ$, including the southern sky, which the Sloan Digital Sky Survey (SDSS) does not cover (see Figure 2). This catalog has optical photometric data in three bands (g, r, z). For 0.17 bn sources, the photometric redshifts and fitting results of various morphological models (e.g., the point-spread function, PSF; exponential, de Vaucouleurs, and composite) are available (Zou et al. 2019).

3.4. WISE

The Wide-field Infrared Survey Explorer (WISE)¹¹ survey covers the entire sky, and the ALLWISE catalog includes 740 million sources with near- or mid-IR photometric

measurements in four bands (3.4, 4.6, 12, and $22 \mu\text{m}$). We used the near- and mid-IR colors to classify stars, galaxies, and QSOs.

3.5. 2MASS

The¹² point-source catalog includes 470 million sources, and the extended source catalog includes 1.6 million sources ($\sim 70,000$ sources are listed only in the extended source catalog). They have photometric measurements in three near-IR bands (j, h, k).

3.6. GALEX

The¹³ revised GUV catalog (Bianchi et al. 2017) includes 83 million sources with UV photometric data in two bands (far-UV, FUV, $\lambda_{\text{eff}} = 1528 \text{ \AA}$, and near-UV, NUV, $\lambda_{\text{eff}} = 2310 \text{ \AA}$).

¹⁰ <https://www.legacysurvey.org>

¹¹ <https://wise2.ipac.caltech.edu/docs/release/allsky/>

¹² <https://irsa.ipac.caltech.edu/data/2MASS/docs/>

¹³ <http://www.galex.caltech.edu>

4. Crossmatch

To crossmatch the CSC2 with our suite of reference catalogs, we used the NWAY package v4.5.2¹⁴ (see Salvato et al. 2018 for details). Given that the CSC2 covers a limited part of the sky (~ 500 deg²), we grouped the stacks¹⁵ into 4380 ensembles, each covering a small part of the continuous sky, each of which we crossmatched individually with the optical or IR catalog data within the same sky footprint.

The NWAY parameters were set to optimize the matching fraction while minimizing false positives and negatives. (a) We started with a search radius of $10''$, considering the reasonable maximum of CSC2 positional errors to avoid losing any real match, but at the end of the match, we restricted this criterion to accept only unique matches with a separation $< 3''$, regardless of the positional error. (b) We set the minimum probability ratio for the secondary match *acceptable-prob* = 0.25—which is lower than the default value of 0.5 - to determine a “unique match” conservatively, particularly in crowded fields with high source densities. (c) We set *match_flag* = 1 (indicating the most probable match). (d) We set the probability threshold (i.e., the probability that one of the associations is correct) to *p_any* = 0.5. With these rather strict selection criteria, we have minimized the false-match rate to $\sim 5\%$, as described below and in Appendix A.

To evaluate the rate of false matches, we have run extensive simulations, following the matching procedure outlined above, but after shifting the source positions in eight directions (horizontal, vertical, and diagonal) by $30''$. Not surprisingly, we found that the rate of false matches is a strong function of source density. When the source density is low (< 10 sources in arcmin²), the rate of false matches is also low ($< 5\%$). When the source density is intermediate (10–30 sources in arcmin²), the rate of false positives is $\sim 10\%$. When the source density is very high (> 40 sources in arcmin²), the rate of false positives can be as high as $\sim 40\%$. We present the complete simulation results in Appendix A.

To minimize the false-match rate by avoiding crowded fields, we limited the range of Galactic latitude ($|b| > 15$) and removed the ensembles with high optical or IR source densities, such as those containing nearby large galaxies (e.g., M31 and M33, and the Large and Small Magellanic Clouds, LMC and SMC). For ensembles partially containing regions of high source density, such as Galactic globular clusters (e.g., M2, M3, and M5) we excised the high-density region. After excluding these crowded fields and applying our strict selection criteria, we achieve a rate of false-positive matches of $\sim 5\%$ in our matches between CSC2 and other catalogs.

An assessment of the quality of the crossmatching can be made based on the spatial separation between the optical or IR counterparts for a given X-ray source. Since the positional uncertainties of the optical or IR catalogs are considerably smaller than those of the CSC2 sources (e.g., $0''.05$ on average for the SDSS), the separation of the matched counterparts should be small if the match is not due to chance. Among the CSC2 sources with multiple optical counterparts from the SDSS, Pan-STARRS, and LEGACY catalogs, we find that the separations between the two optical sources exceed $1''$ for only 0.4% of the matches. The fraction becomes higher (3%) when

Table 1
SDSS Counterparts

| Match Condition | Number of Objects |
|--|---------------------------|
| Total number of matches | 34,085 |
| Objects with high-quality spec-z | 9350 ^a |
| Objects with high-quality photo-z (without spec-z) | 7617 (4834 ^b) |
| Objects with morphological type from Galaxy Zoo | 1249 |

Note.

^a SDSS spec-z sample.

^b SDSS photo-z sample.

2MASS counterparts are included and is highest (9%) when WISE counterparts are included. Based on this, we further excluded from the CGC all the CSC2 sources with a large separation between two counterparts ($> 1''$ without WISE or $> 2''$ with WISE). This additional condition removes $\sim 1\%$ of the matches for which at least one counterpart is likely false.

5. Classification

5.1. SDSS Classification and Sample Extraction

Crossmatching CSC2 and SDSS, we found $\sim 34,000$ unique SDSS counterparts (Figure 2). We summarize the number of objects with spectroscopic, photometric, and morphological data in Table 1. From these matches, we have defined two high-quality subsamples, the spec-z and the photo-z samples, as described below. A detailed justification of our selection criteria is given in Appendix A.

To assemble the spec-z sample, we have used SDSS counterparts with spectroscopic information with *zWarning* = 0, to exclude low-quality spectroscopic redshifts. We also excluded a small number of objects with large errors (*error(z)/z* > 0.5 ; see Bolton et al. 2012 for the list of SDSS quality checks). After removing ~ 600 matches with low-quality data, we have 9350 objects with high-quality spec-z and spectral classes.

For the photo-z sample, we use the photometric redshifts (photo-z) from Beck et al. (2016).

They used multiple training sets from optical galaxy catalogs and applied various galaxy templates in spectral energy distribution (SED) fitting. As their sample effectively excluded quasi-stellar objects (QSOs) and stars, our photo-z sample primarily consists of galaxies (see Section 5.1.3). From this sample, we selected only objects with *photoErrorClass* = 1 (Beck et al. 2016). We also exclude a small number of objects with *error(z)/z* > 0.5 . This selection results in the removal of $\sim 60\%$ of the photo-z SDSS counterparts, leaving 7617 objects with high-quality photo-z. Of these, 4834 are not included in the SDSS spec-z sample.

The spectroscopic classification of the spec-z sample identifies 22% of the sample as galaxies, 77% as QSOs, and 0.4% as stars (see Table 2). Since we cannot determine the distances of stellar CSC2 sources in the Milky Way from the SDSS information, we will exclude the CSC2 sources with “star” counterpart from most of the considerations of this paper, but we will consider this sample for future work.

Of the galaxies, 23% show star formation and $\sim 60\%$ are classified as unknown (NaN) based on the SDSS subclass.¹⁶ The latter likely includes a large number of early-type galaxies

¹⁴ <https://github.com/JohannesBuchner/nway>

¹⁵ <https://cxc.cfa.harvard.edu/csc/about.html>

¹⁶ <https://www.sdss.org/dr16/spectro/catalogs/#Objectinformation>

Table 2
SDSS Spectroscopic Subclass

| Subclass | Galaxy | QSO | Star |
|--------------|--------|------|------|
| Total | 2071 | 7240 | 39 |
| Broad-line | 93 | 5571 | |
| Starburst | 125 | 316 | |
| Star-forming | 349 | 23 | |
| AGN | 293 | 80 | |
| NaN | 1211 | 1255 | |

with no detectable emission lines. The remaining 18% are assigned as subclass broadline or AGN. We checked whether these objects with some AGN signatures differ from the other spectroscopically classified galaxies in their L_X , F_{XO} , and WISE colors and found no significant difference (see Figures 2 and 13 in Kim et al. 2023). Since we primarily rely on L_X , F_{XO} , and WISE colors to select galaxies, we consider the entire 2071 objects with class = galaxy as the spectroscopically classified galaxies. We also include the subclass in our catalog so that users can apply their own selections for their specific goals. We note that we will continue to investigate multiwavelength properties (from radio to X-ray) in different types of galaxies (early type, late type, AGN, LINERS, etc.) in forthcoming papers.

Among the spectroscopically classified QSOs, three-quarters exhibit broad emission lines, while 5% are star forming and 17% are unknown. The spectroscopically classified stars also have subclasses (e.g., stellar spectral types), but we do not list them here.

5.1.1. X-Ray-to-optical Flux Ratio ($F_{XO} = F_X/F_o$) Distributions

We define the X-ray-to-optical flux ratio F_{XO} as in Maccacaro et al. (1988): $\log(F_{XO}) = \log(F_X) + 5.31 + r/2.5$, and calculated F_{XO} values for the matched sources from the broadband (0.5–7 keV) X-ray flux from CSC2 and the r -band magnitude in the SDSS *PhotoObj* table. F_{XO} has been used in the literature to discriminate between galaxies and AGNs (e.g., Maccacaro et al. 1988, Kim et al. 1992, Kim et al. 2006; Shapley et al. 2001). Most AGNs have F_{XO} values within the range of 0.1–10, while galaxies dominate at lower F_{XO} (<0.1).

In Figure 3 (left) we plot the r magnitude versus the X-ray flux (F_X) in the spec- z sample. The points are color-coded to show the SDSS spectroscopic class. We overplot lines representing three different F_{XO} values: 0.01, 0.1, and 10. At lower F_{XO} (<0.01), the distribution is overwhelmingly dominated by galaxies (red points), but objects spectroscopically classified as galaxies are also found at higher values of F_{XO} , overlapping the distribution of QSOs (blue points).

In Figure 3 (right) we show the histogram of F_{XO} for galaxies, QSOs, and stars. QSOs fall within the F_{XO} range of 0.1–10, as expected. Galaxies are in the F_{XO} range of 10^{-4} –10. Stars are mixed with galaxies at F_{XO} (0.001–1). Typical late-type stars (K, M) have F_{XO} in this range (e.g., Zombeck 1990).

Table 3 lists the fraction of each class in different F_{XO} bins. In the lowest F_{XO} bin ($F_{XO} < 0.01$), most sources (94%) are spectroscopically classified galaxies; there are 671 galaxies in this bin, about one-third of 2071 spectroscopically classified galaxies. In the intermediate F_{XO} bin (0.01–0.1), galaxies are the majority (60%), but QSOs account for one-third of the

objects. In the highest F_{XO} bin ($F_{XO} > 0.1$), QSOs dominate ($\sim 90\%$).

We estimated the number of galaxies we expect in the photo- z sample, assuming the same galaxy fraction as found in the spec- z sample. These estimates are listed in the last two columns of Table 3. We expect ~ 1204 galaxies in the photo- z sample. However, there are only 244 galaxies in the lowest F_{XO} bin, where we can identify galaxies with a high confidence of 97% (see Table 3).

5.1.2. Redshift and X-Ray Luminosity (L_X) Distributions

In Figure 4 we show the redshift and L_X distributions of each class in the spec- z sample. Most galaxies in our sample are within the redshift range $z = 0.01$ –1, while most QSOs are at higher redshifts, $z = 0.1$ –10. QSOs dominate at $L_X > 10^{43}$ erg s^{-1} . At $L_X < 10^{43}$ erg s^{-1} , galaxies are most abundant.

Table 4 lists the fraction of each spectroscopically classified object (galaxy and QSO) in the spec- z sample in two L_X bins. In the low L_X bin ($<10^{42}$), galaxies are the majority (98%). This is consistent with what was shown by observations of the near Universe. For this reason, $L_X = 10^{42}$ erg s^{-1} has been used as a discriminator between galaxies and AGNs (e.g., Moran et al. 1999; Kim et al. 2019). However, only 45% of the spectroscopically classified galaxies (945 out of 2071) are found in this L_X bin, as also shown in Figure 4. In the high L_X bin ($L_X > 10^{42}$), QSOs dominate ($\sim 87\%$), but this bin also contains 55% of the CSC2 sources with galaxy spectral classification.

We estimated the number of galaxies we can expect from the photo- z sample based on the galaxy fraction in the spec- z sample (see the last two columns in Table 4). In total, we expect to have about 1292 galaxies from the photo- z sample. However, there are only 742 galaxies in the low L_X bin, where we can identify galaxies with a high confidence of 98% (see Table 4).

For completeness, we describe in Appendix F the redshift and L_X distributions of different samples, including the spec- z sample from SDSS and the photo- z samples from SDSS, Pan-STARRS, and Legacy. They do not vary significantly from one sample to the next.

5.1.3. Using F_{XO} and L_X Together as Classifiers

Here we show that using both F_{XO} and L_X simultaneously provides a more efficient way to identify galaxies. This method can be applied to the SDSS photo- z sample and to the Pan-STARRS and Legacy catalog crossmatches, where no spectroscopic classification is available (see Sections 7 and 8). Figure 5 (top panel) shows the distribution of the spec- z sample in the F_{XO} – L_X plane. As in Figure 3, QSOs and galaxies are color coded. It is obvious that the two classes are better separated in the L_X – F_{XO} plane than by means of L_X or F_{XO} alone. The spectroscopically classified galaxies (red points in Figure 5) lie in a narrow strip with a slope of ~ 1 in the L_X – F_{XO} plane. The best-fit relation (the red line in the top panel of Figure 5) gives $(L_X/2.8 \times 10^{42} \text{ erg } s^{-1}) = (F_{XO}/0.1)^{1.02}$ for galaxies in the spec- z sample. Throughout this paper, we use a Bayesian approach to linear regression (Kelly 2007). The linear L_X – F_{XO} relation runs for about five orders of magnitude in both L_X (10^{39} – 10^{44}) and F_{XO} (10^{-4} –10). The width of the strip is narrow, narrower than 1 dex, regardless of L_X . This is primarily because the dynamic range of the optical stellar luminosity (L_r)

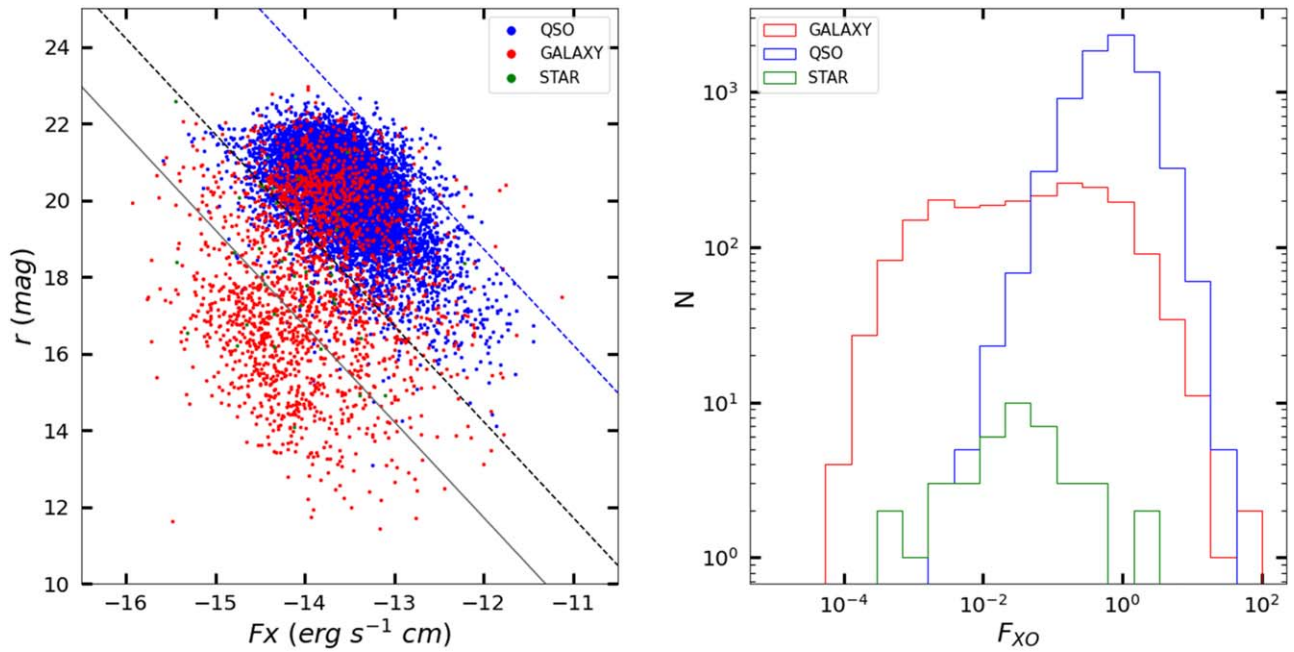


Figure 3. (Left) Optical r mag against X-ray flux for the SDSS spec- z sample. The diagonal lines indicate F_{XO} values of 0.01, 0.1, and 10 from the lower left to the upper right. (Right) F_{XO} distribution for the QSO, galaxy, and star subsamples.

Table 3
Number of Objects in Different F_{XO} Bins for the SDSS Samples

| F_{XO} | Spec- z Sample | | | | Photo- z Sample | |
|----------|------------------|------------|---------|-------|-------------------|-----------------------|
| | Galaxies | QSOs | Stars | Total | Total | Galaxies ^a |
| <0.01 | 671 (97%) | 10 (2%) | 10 (1%) | 691 | 252 | 244 ^b |
| 0.01–0.1 | 533 (61%) | 323 (37%) | 21 (2%) | 877 | 905 | 550 |
| >0.1 | 867 (11%) | 6907 (89%) | 8 (0%) | 7782 | 3677 | 410 |
| Total | 2071 | 7240 | 39 | 9350 | 4834 | 1204 |

Notes.

^a Expected number of galaxies in each F_{XO} bin from the photo- z sample.

^b Number of galaxies without significant contamination.

is narrow, with 90% of galaxies having L_r in a range of 20 (10^{10} – $2 \times 10^{11} L_{r\odot}$; see the bottom panel of Figure 5), to be compared with a range of 10^4 in L_X (10^{40} – 10^{44} erg s⁻¹).¹⁷ While the upper limit of L_r is likely intrinsic, the lower limit is partly due to the selection effect that small galaxies are undetected in X-rays unless they are nearby. Nonetheless, the extensive dynamical range of L_X means that L_X is practically independent of L_r in our sample.

In contrast to the galaxy sample, QSOs are clustered in the upper right corner in the F_{XO} – L_X plane with the highest L_X and F_{XO} . There is no correlation between L_X and F_{XO} in the QSO sample. Instead, QSOs show a correlation between L_X and L_r (bottom right panel of Figure 5). This can be understood because the strong nuclei enhance their fluxes both in the optical and X-ray bands. The best-fit relation (the blue line in the bottom panel of Figure 5) is given by $(L_X / 1.6 \times 10^{43}) = (L_r / 10^{10})^{0.72}$ for QSOs in the spec- z sample. This relation is

¹⁷ We determine the r -band luminosity (L_r) in units of solar luminosity, assuming an absolute solar magnitude $M_{r\odot} = 4.68$ mag.

close to the well-known L_X – L_{UV} correlation of (unabsorbed) QSOs with a slope of 0.6 ± 0.1 (Tananbaum et al. 1979).

To identify galaxies with the least contaminations by QSOs, we empirically draw a boundary line in the L_X – F_{XO} plane, marked by the blue line in the top panel of Figure 5, $(L_X / 4.64 \times 10^{42}) = (F_{XO} / 0.1)^{-1/3}$ the galaxy–QSO boundary.

In Table 5 we list the number of galaxies in the two regions (1) above and (2) below the galaxy–QSO boundary line. From region 2, we recover 1311 galaxies (63%) out of 2071 spectroscopically classified galaxies. This fraction is considerably higher than that by F_{XO} alone (one-third; see Section 5.1.1) and that by L_X alone (half; see Section 5.1.2). The QSO contamination in this region is 6%.

The missing galaxies (37% of the total spectroscopically classified galaxies) are in the high L_X and F_{XO} region. These X-ray bright galaxies have been called in the literature “X-ray bright optically normal galaxies” (XBONG; e.g., Elvis et al. 1981; Fiore et al. 2000). We will explore this unusual type of object in a separate paper (Kim et al. 2023)

Figure 6 compares the distributions of photo- z and spec- z samples in the L_X – F_{XO} plane (the top panel) and the L_X – L_r plane (the bottom panel). The distribution of the photo- z sample (the red points in Figure 6; as discussed in Section 5.1, this sample primarily consists of galaxies) has a virtually identical slope to that of the spectroscopically classified galaxies (the red points in Figure 5), but a 0.1 dex lower intercept. The best-fit relation of the photo- z sample is $(L_X / 2.2 \times 10^{42}) = (F_{XO} / 0.1)^{1.01}$ for the photo- z sample and extends to the XBONG regime, as observed in the spec- z sample.

As in Sections 5.1.1 and 5.1.2, we also estimate the number of galaxies expected from the photo- z sample (see the last two columns in Table 5). From the photo- z sample, the total number of objects in region 2 between the two boundary lines is 1642. We expect 1540 galaxies with a contamination fraction of 6%. However, the contamination is expected to be lower because the photo- z sample preferentially includes galaxies.

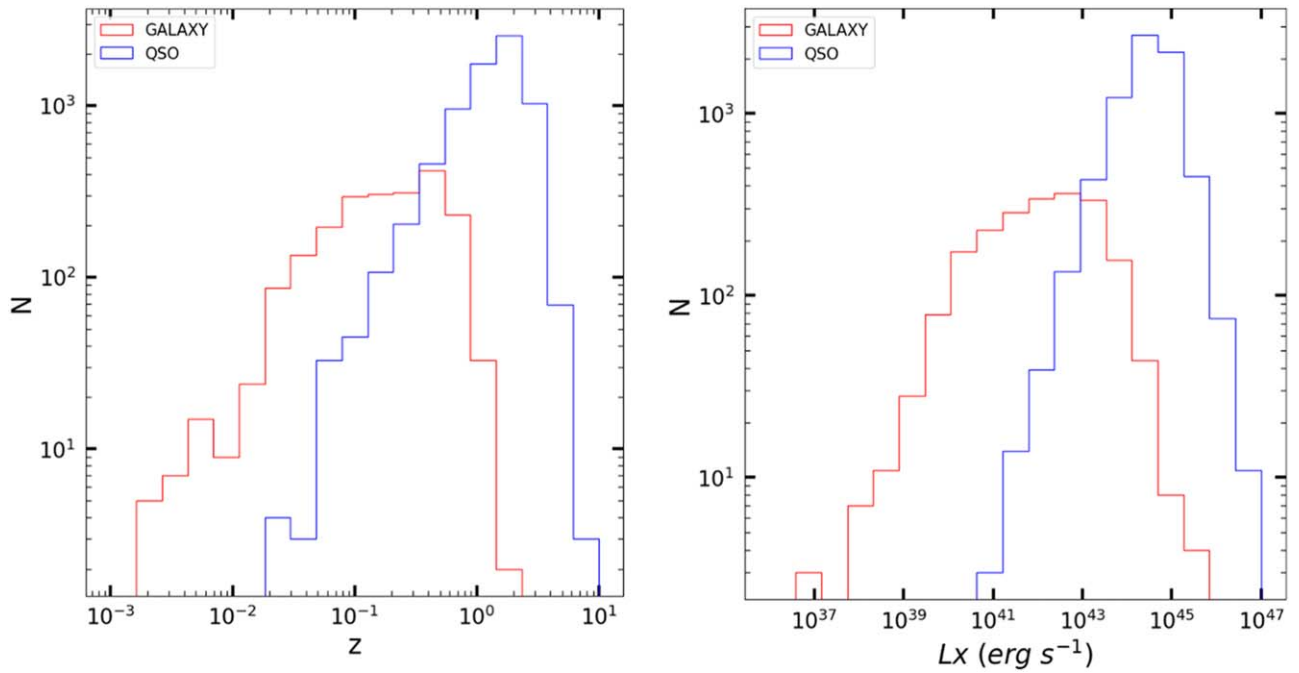


Figure 4. The redshift and L_X distributions of galaxies and QSOs for the SDSS spec- z sample.

Table 4
Number of Objects in Different L_X Bins for the SDSS Samples

| L_X (erg s $^{-1}$) | Spec- z Sample | | | Photo- z Sample | |
|------------------------|------------------|------------|-------|-------------------|-----------------------|
| | Galaxies | QSOs | Total | Total | Galaxies ^a |
| <1e42 | 945 (98%) | 24 (2%) | 969 | 757 | 742 ^b |
| >1e42 | 1126 (13%) | 7216 (87%) | 8342 | 4077 | 550 |
| Total | 2071 | 7240 | 9311 | 4834 | 1292 |

Notes.

^a Expected number of galaxies in each L_X bin from the photo- z sample.

^b Number of galaxies without significant contamination.

In summary, we find 2071 galaxies from the SDSS spec- z sample and identify 1642 galaxy candidates from the SDSS photo- z sample. In total, we obtain an X-ray galaxy catalog with 3713 galaxy candidates with 102 (= 1642 – 1540) nongalaxies, i.e., a contamination fraction of 3%.

5.2. CSC2-SDSS-WISE Sample Selection

In addition to F_{XO} and L_X , we have used the WISE colors to identify galaxies and exclude QSOs. Different types of objects occupy unique positions in the IR color-color space (e.g., Wright et al. 2010; Jarrett et al. 2017). We have crossmatched the CSC2 and WISE (All-WISE) catalogs with the same method described in Section 4, finding 124,889 matches. Because the optical information is necessary for the optical magnitude and redshift, from these matches, we selected 85,314 CSC-WISE matches with optical counterparts in the SDSS, Pan-STARRS, and Legacy catalogs. We then applied a set of quality-control criteria and identify 25,190 WISE counterparts. These criteria include limiting Galactic latitude ($|b| > 15$ as in Sections 2.1 and 4), excluding well-known crowded fields (Sections 2.1 and 4), limiting the separations between optical and IR counterparts ($< 2''$; Section 4), and removing objects with low-quality z data (Section 5.1). Furthermore, we selected sources with a signal-to-

noise ratio $S/N > 2$ in two and three of the WISE bands, W1 (3.4 μm), W2 (4.6 μm), and W3 (12 μm), as listed in Table 6. We have 13,627 WISE-CSC matches with $S/N > 2$ in three bands. Of these, 7956 have redshifts from SDSS, 5482 with spec- z , and 2474 with photo- z . We call them the WISE spec- z and photo- z samples, respectively.

5.2.1. WISE IR Color Classification

Figure 7 plots the WISE spec- z and photo- z samples in the W_{12} – W_{23} color-color diagram, where $W_{12} = W1$ – $W2$ and $W_{23} = W2$ – $W3$. The IR magnitudes are in the Vega magnitude system. The top panel of Figure 7 shows the WISE spec- z sample (5480 objects)—spectroscopically classified galaxies and QSOs color-coded, as in Figure 5. QSOs are clustered at higher W_{12} (0.5–1.7) and W_{23} (2–4) colors. In contrast, galaxies are at lower W_{12} colors (0–1), but in a wide range of W_{23} (0–4.5). Previous AGN surveys often applied a lower limit of W_{12} to identify QSOs. For example, Jarrett et al. (2011) and Stern et al. (2012) applied $W_{12} > 0.8$, while Assef et al. (2013) applied a similar but W_2 -dependent limit. The dashed blue line in Figure 7 indicates $W_{12} = 0.8$. As seen in the figure, this limit is helpful in selecting QSOs with a small galaxy contamination (3%). Under this limit ($W_{12} < 0.8$), galaxies dominate, but the QSO fraction is still high. To effectively select galaxies by reducing the QSO contamination, we apply a strict upper limit of $W_{12} < 0.4$ for galaxies. The number of objects in different W_{12} bins is summarized in Table 7. In the lowest W_{12} bin (< 0.4), 93% of objects are galaxies with 7% contamination by QSOs. In the middle bin ($W_{12} = 0.4$ – 0.8), galaxies and QSOs are mixed roughly by a ratio of 1:2.

In the bottom panel of Figure 7, the WISE spec- z and photo- z samples are compared, color coded as in the bottom panel of Figure 6. In contrast to the F_{XO} – L_X plot in Figure 6, where the distribution of the photo- z sample resembles that of the spec- z galaxy sample, the WISE photo- z sample falls in between the spec- z galaxy sample and QSO samples, indicating that galaxies and QSOs are mixed in the WISE photo- z sample.

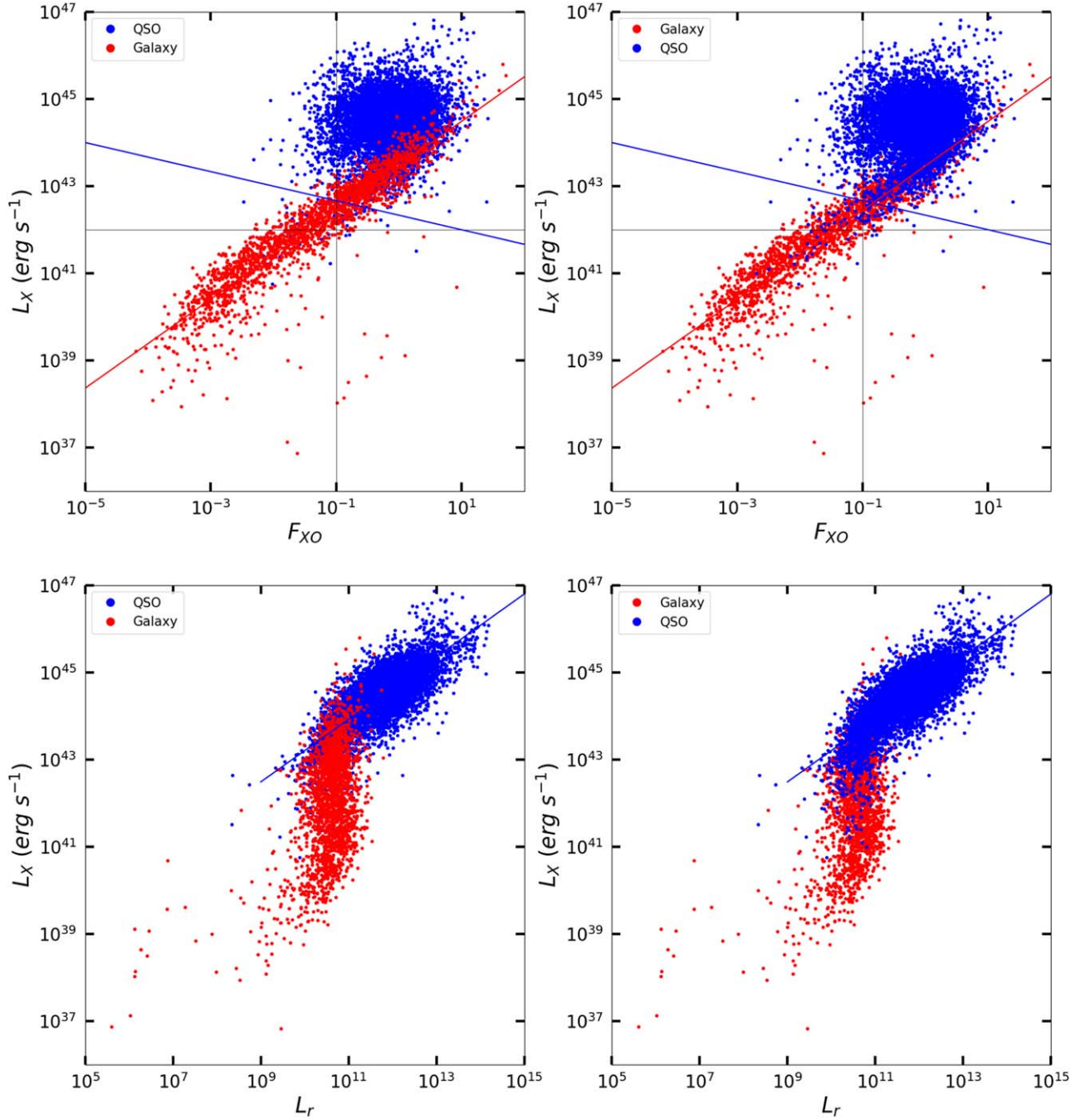


Figure 5. (Top) L_X against F_{XO} for galaxies and QSOs from the spec- z sample. The left and right panels are the same, except for (left) galaxies on top of QSOs or (right) QSOs on top of galaxies to visualize both subsamples clearly. The dashed red line is a linear relation that fits the galaxy sample. The blue line separates galaxies from QSOs. (Bottom) The same, but with L_X against L_r . The blue line is the best-fit relation for QSOs.

Table 5

Number of Objects in Different F_{XO} and L_X Bins for the SDSS Samples

| (F_{XO}, L_X) | Spec- z Sample | | | Photo- z Sample | |
|-----------------|------------------|------------|-------|-------------------|-----------------------|
| | Galaxies | QSOs | Total | Total | Galaxies ^a |
| Region 1 | 760 (10%) | 7151 (90%) | 7911 | 3192 | 307 |
| Region 2 | 1311 (94%) | 89 (6%) | 1400 | 1642 | 1540 ^b |
| Total | 2071 (22%) | 7240 (78%) | 9311 | 4834 | 1847 |

Notes. Region 1: above the galaxy–QSO boundary line. Region 2: below the galaxy–QSO boundary line.

^a Expected number of galaxies in each (F_{XO}, L_X) bin from the photo- z sample.

^b Number of galaxies without significant contamination.

Assuming the same galaxy fraction in each W_{12} bin, we expect that 46% (1142 out of 2474) of the WISE photo- z sample will be galaxies (see the last two columns in Table 7). In the first bin ($W_{12} < 0.4$), where galaxies dominate, we can identify 760 galaxy candidates with 7% QSO contamination.

5.2.2. L_X – W_{12} Classification

While W_{23} is a good indicator to separate passive galaxies with low W_{23} colors and SF galaxies with high W_{23} colors (e.g., Wright et al. 2010; see also Section 6), there is considerable overlap of galaxies and QSOs in W_{23} color (Figure 7). Therefore, we used W_{12} together with L_X to explore

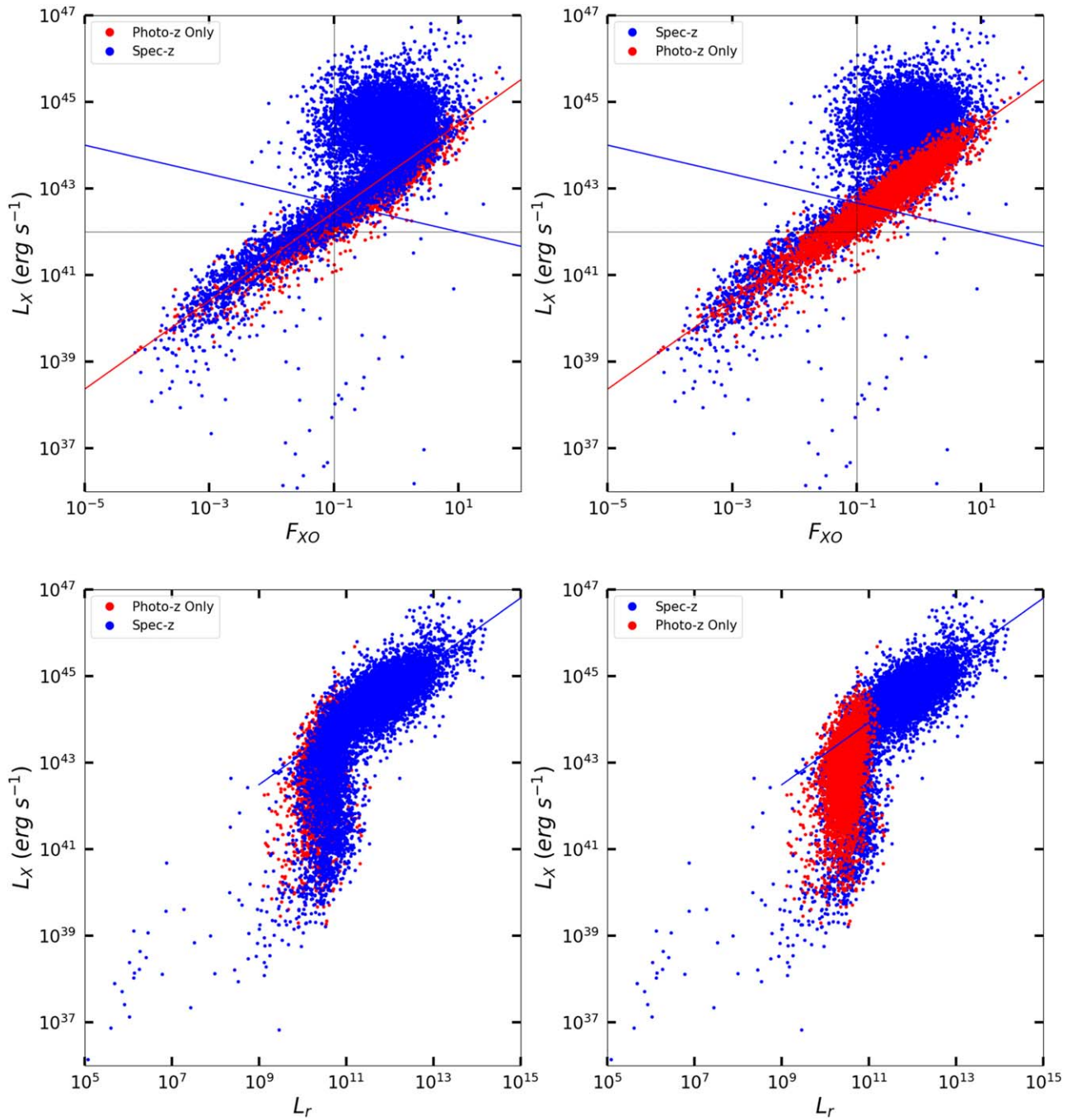


Figure 6. L_X against (top) F_{XO} and (bottom) L_r from spec-z and photo-z samples. The left and right panels are the same, except for (left) the spec-z sample on top of the photo-z sample or (right) vice versa to visualize both samples clearly. The diagonal lines are the same as in Figure 5.

Table 6
CSC–SDSS–WISE Matches

| CSC–WISE | All | Selected ^a | SDSS Spec-z | SDSS Photo-z (without Spec-z) |
|--------------------|--------|-----------------------|-------------------|----------------------------------|
| WISE detected | 85,314 | 25,190 | 7991 | 4396 |
| S/N (W1,W2) > 2 | 82,181 | 24,684 | 7920 | 4355 |
| S/N (W1,W2,W3) > 2 | 41,177 | 13,627 | 5482 ^b | 2474 ^c |

Notes.

^a Selected e.g., by Galactic latitude, z quality, and crowd fields (see text).

^b WISE spec-z sample.

^c WISE photo-z sample.

Table 7
Number of Objects in Different W₁₂ (W1–W2) Bins for the SDSS Samples

| W12 (W1–W2) | Spec-z Sample | | | Photo-z Sample | |
|-------------|---------------|------------|-------|----------------|-----------------------|
| | Galaxies | QSOs | Total | Total | Galaxies ^a |
| <0.4 | 858 (93%) | 68 (7%) | 926 | 760 | 704 ^b |
| 0.4–0.8 | 382 (35%) | 720 (65%) | 1102 | 1245 | 431 |
| >0.8 | 90 (3%) | 3352 (97%) | 3442 | 469 | 12 |
| Total | 1330 (24%) | 4140 (76%) | 5470 | 2474 | 1142 |

Notes.

^a Expected number of galaxies in each W12 bin from the photo-z sample

^b Number of galaxies without significant contamination.

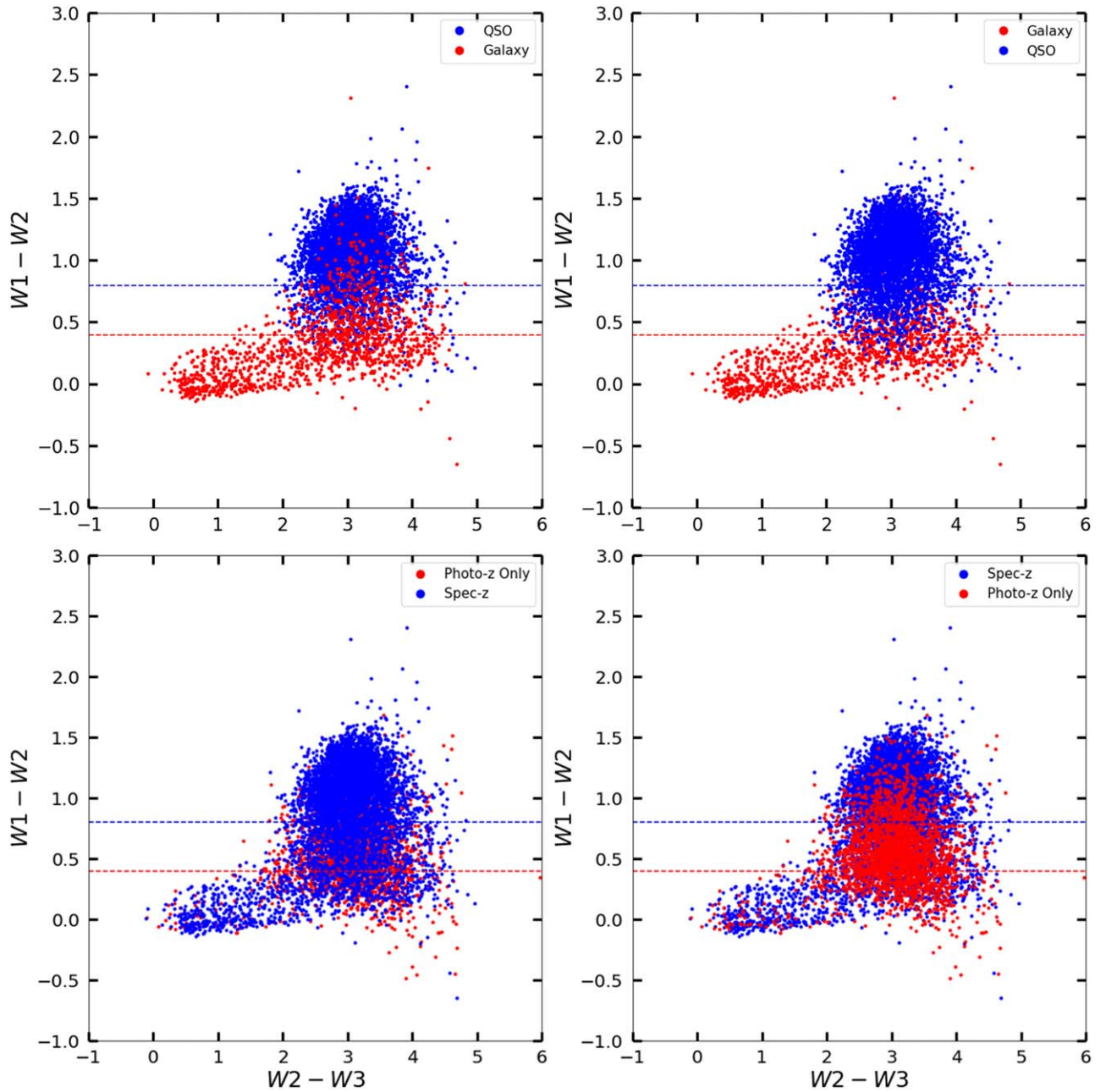


Figure 7. Top: WISE spec- z color-color plot for galaxies and QSOs. Bottom: comparison of WISE spec- z and photo- z samples. The left and right panels are the same, but the different objects are drawn on top to visualize both samples clearly. The dashed blue line is for $W_{12} = 0.8$, often used to identify QSOs without galaxy contaminations. The dashed red line is for $W_{12} = 0.4$, which we apply to select galaxies without QSO contaminations.

a more efficient way to separate galaxies, QSOs, and sources with stellar SDSS counterparts (Figure 8). We selected a boundary line in the L_X - W_{12} plane to separate galaxies from QSOs (shown in Figure 8 by the blue line), $\log(L_X) = -2.22 W_{12} + 43.78$ the galaxy-QSO boundary.

In Table 8 we list the number of galaxies in the two regions (1) above and (2) below the galaxy-QSO boundary line. From region 2, we recover 954 galaxies (72%) out of 1330 spectroscopically classified galaxies. The nongalaxy contaminations in these regions are only 4%, all being QSOs. When compared to the efficiency of finding galaxies in the L_X - F_{XO} plane (Figure 6), the L_X - W_{12} plane provides a slightly more effective result (72% versus 64%). The contamination fraction is also slightly improved (4% versus 6%).

The missing galaxies (28% of the spectroscopically classified galaxies) are found in region 1, with the highest L_X and W_{12} . These X-ray bright galaxies correspond to XBONGs; we will explore them in a separate paper (Kim et al. 2023).

In the bottom panel of Figure 8, the WISE photo- z sample (2474 objects) is compared with the WISE spec- z sample (5482 objects). The last two columns in Table 8 show the number of galaxies expected in the photo- z sample, based on the spectroscopically classified galaxies. In the photo- z sample, the total number of objects in region 2 between the two boundary lines is 944. We expect 902 galaxies with a contamination fraction of 4%, as estimated from the WISE spec- z sample.

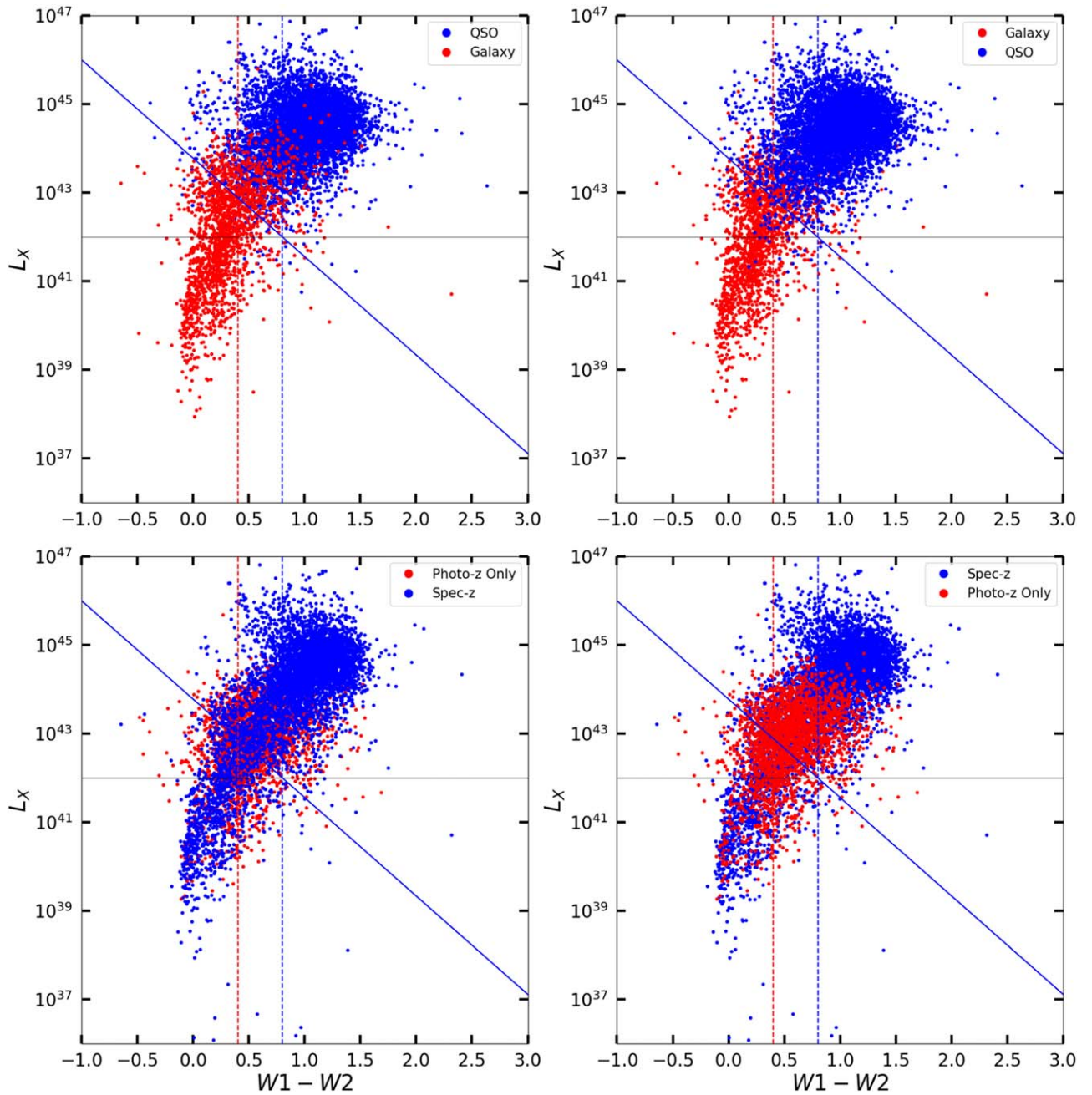


Figure 8. L_X against WISE color W_{12} ($= W1 - W2$) (top) for galaxies and QSOs from the SDSS spec-z sample (bottom) from the SDSS spec-z and photo-z samples. The left and right panels are the same, but the different objects are drawn on top to visualize both samples clearly. The blue line separates galaxies and QSOs.

Table 8

Number of Objects in Different W_{12} and L_X Bins for the SDSS Samples

| W1-W2 | Spec-z Sample | | | Photo-z Sample | |
|----------|---------------|------------|-------|----------------|-----------------------|
| | Galaxies | QSOs | Total | Total | Galaxies ^a |
| Region 1 | 376 (8%) | 4096 (92%) | 4472 | 1530 | 129 |
| Region 2 | 954 (96%) | 44 (4%) | 998 | 944 | 902 ^b |
| Total | 1330 (24%) | 4140 (76%) | 5470 | 2474 | 1031 |

Notes. Region 1: above the galaxy–QSO boundary line. Region 2: below the galaxy–QSO boundary line.

^a Expected number of galaxies in each W_{12} bin from the photo-z sample.

^b Number of galaxy candidates without significant contamination.

In summary, we find 2071 Chandra-detected galaxies in the SDSS spec-z sample. Additionally, in the SDSS photo-z sample, we identify 1642 and 944 galaxies candidates using the $L_X - F_{XO}$ selection and the $L_X - W_{12}$ selection, respectively; 734 galaxies are found by both selection methods. In total, we obtain an X-ray galaxy catalog with 3939 galaxy candidates with a contamination fraction of 3%.

5.3. Pan-STARRS and Legacy Crossmatch Samples

Given that the sky coverage of SDSS is limited, we also used two additional wide-area optical survey catalogs:

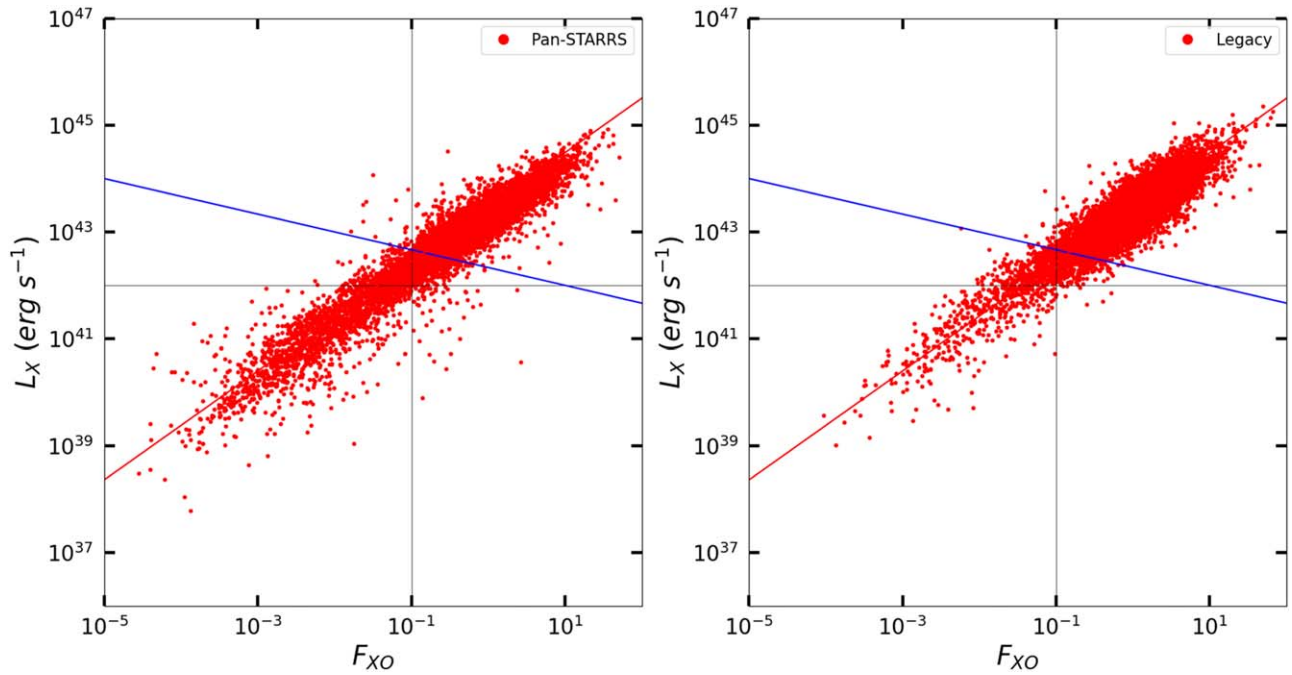


Figure 9. L_X against F_{XO} for (left) the PS and (right) the LS samples. All lines are the same as in the top panel of Figure 6.

1. Pan-STARRS (Flewelling et al. 2020), which covers three-quarters of the entire sky ($\delta > -30^\circ$; see Figure 2 for the sky maps). In the following, we use the abbreviation PS to refer to the Pan-STARRS DR2 catalog. Since PS has no spectroscopic information, we use the photo- z from Beck et al. (2021), who applied a machine-learning neural network algorithm trained with a spectroscopic sample of redshifts and classification information to determine a probabilistic photometric classification: Galaxy, star, QSO, or unsure, and photometric redshifts for sources identified as galaxies. In Appendix B we show that these photo- z estimates are consistent with the SDSS spec- z and photo- z . For the r magnitude, we take the Kron magnitude because it is suitable for measuring the total flux of an extended source, such as a galaxy (Kron 1980). If not available, we take the AP magnitude (Flewelling et al. 2020). We validate these choices by comparing r magnitudes for the objects common in SDSS and PS (Appendix C).
2. The DESI Legacy catalog (Dey et al. 2019). The Legacy South covers the sky area outside the SDSS footprint, while the Legacy North footprint significantly overlaps with that of SDSS (see Figure 1). In this work, we only use the Legacy South data. We use the abbreviation LS to refer to the Legacy South DR8 catalog. We take the photo- z from Zou et al. (2019). In Appendices B and C, we show that the LS photo- z and r mag are consistent with those of SDSS and PS.

Applying the crossmatch procedure described in Section 4, we find 105,728 (69,478) PS (LS) counterparts to the CSC2 sources (Table 9), which are reduced to 22,617 (20,204) PS (LS) objects with the set of quality selection criteria described in Section 5.2.

Further excluding those already matched with SDSS objects, we find 8762 PS counterparts: the “PS sample”. Of these, 3396

Table 9
Pan-STARRS and Legacy Counterparts

| | Pan-STARRS | Legacy |
|-------------------------|-------------------|-------------------|
| Total number of matches | 105,728 | 69,478 |
| Selected ^a | 22,617 | 20,204 |
| Used here | 8762 ^b | 7642 ^c |
| With WISE detection | 3396 ^d | 2259 ^e |

Notes.

- ^a Selections by Galactic latitude, z quality, and crowd fields (see text).
- ^b PS sample (do not have SDSS counterparts).
- ^c LS sample (do not have SDSS and Pan-STARRS counterparts).
- ^d PS-WISE sample ($S/N > 2$ in W1, W2, and W3).
- ^e LS-WISE sample ($S/N > 2$ in W1, W2, and W3).

are also detected with $S/N > 2$ in the three WISE bands W1, W2, and W3, which we call the “PS-WISE sample”.

Similarly, further excluding those already matched with SDSS and PS objects, we find 7642 LS counterparts, constituting the “LS sample”. Of these, 2259 are detected with $S/N > 2$ in the WISE three bands W1, W2, and W3: the “LS-WISE” sample.

All the objects in the PS and LS samples are photometrically classified as galaxies by Beck et al. (2021) and Zou et al. (2019), who apply galaxy templates to derive photometric redshifts. Figure 9 shows the PS and LS samples in the L_X – F_{XO} plane. Note that their distributions are similar to those of the SDSS galaxies (the red points in Figure 5). The diagonal red line is the best fit (with slope 1.02) from the SDSS spec- z galaxy sample. This line represents the PS and LS samples well. The best slopes for the PS and LS samples are 0.98 and 0.96, respectively.

The distributions of the PS-WISE and LS-WISE samples in the W_{12} – W_{23} and L_X – W_{12} planes (Figures 10 and 11) are similar to those of the SDSS photo- z sample. In Tables 10 and 11, we list the number of objects in three regions of the

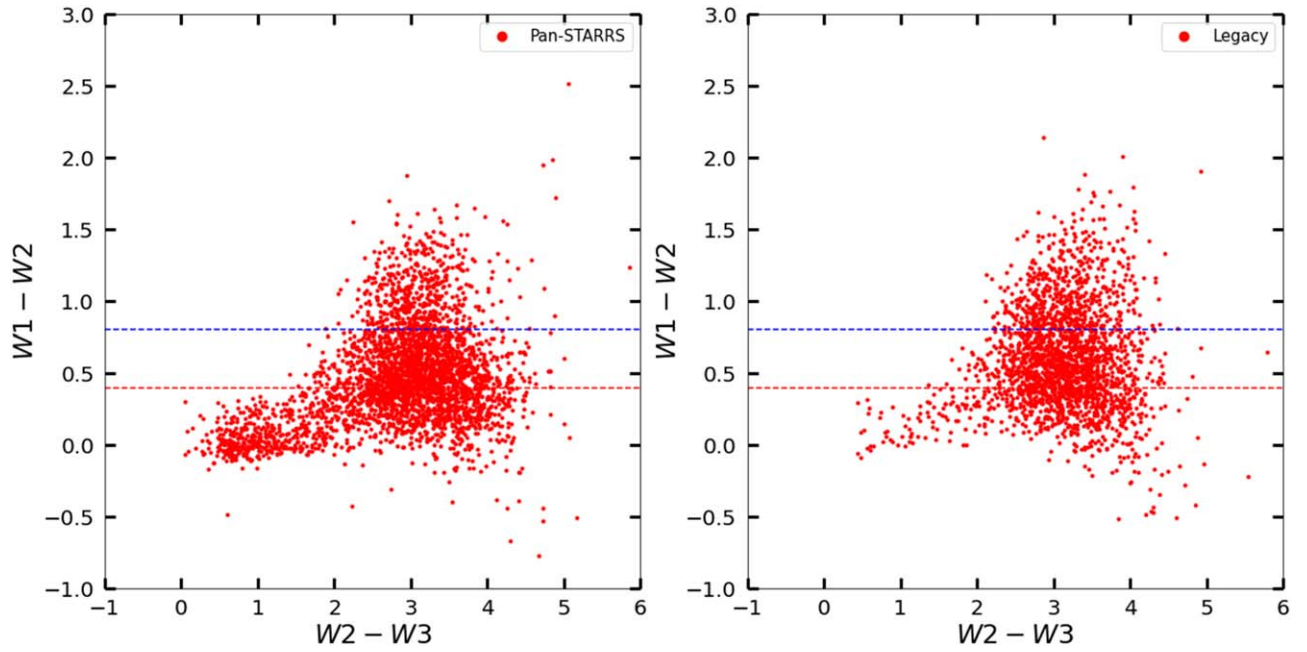


Figure 10. W_{12} ($W1-W2$) against W_{23} ($W2-W3$) for (left) the PS-WISE sample and (right) the LS-WISE sample. All lines are the same as in Figure 7.

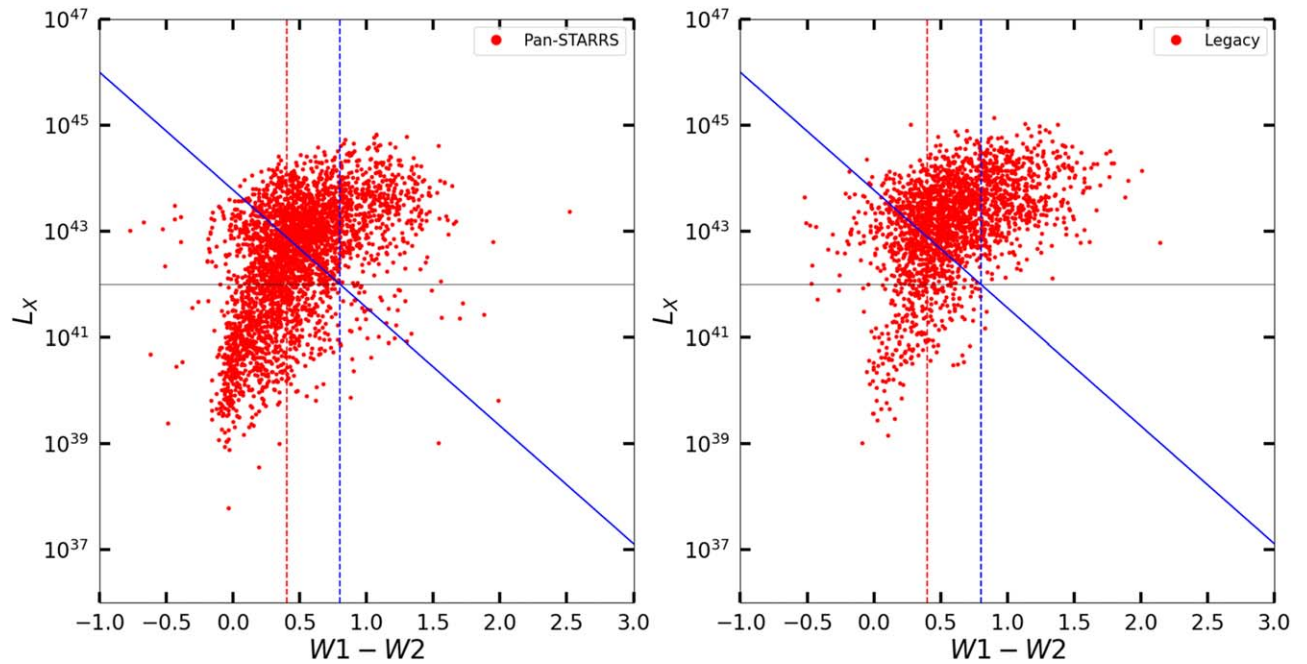


Figure 11. L_X against W_{12} for (left) the PS-WISE sample and (right) the LS-WISE sample. All lines are the same as in Figure 8.

(L_X-F_{XO}) and (L_X-W) planes. Applying the L_X-F_{XO} selection, we identify 3060 and 1121 (total 4181) galaxy candidates from the PS and LS samples, respectively. Applying the L_X-W_{12} selection, we find 1787 and 653 (total 2440) galaxy candidates from the PS and LS samples, respectively. Of these objects, 1455 and 438 (total 1893) satisfy both selection criteria. In total, we find 4708 galaxy candidates with a contamination fraction of $\sim 5\%$.

In summary, we find 2071 Chandra-detected galaxies from the SDSS spec- z sample. Additionally, from the SDSS/PS/LS photo- z sample, we identify 5823 and 3384 galaxy candidates

Table 10
Number of Objects in Different F_{XO} and L_X Bins for the Pan-STARRS and Legacy Samples

| (F_{XO}, L_X) | Pan-STARRS | Legacy |
|-----------------|-------------------|-------------------|
| Region 1 | 5702 | 6519 |
| Region 2 | 3060 ^a | 1121 ^a |
| Total | 8762 | 7640 |

Notes. Region 1: above the galaxy-QSO boundary line. Region 2: below the galaxy-QSO boundary line.

^a Number of galaxy candidates without significant contamination.

Table 11Number of Objects in Different W_{12} and L_X Bins for the Pan-STARRS and Legacy Samples

| (W_{12} , L_X) | Pan-STARRS | Legacy |
|----------------------|-------------------|------------------|
| Region 1 | 1609 | 1606 |
| Region 2 | 1787 ^a | 653 ^a |
| Total | 3396 | 2259 |

Notes. Region 1: above the galaxy–QSO boundary line. Region 2: below the galaxy–QSO boundary line.

^a Number of galaxy candidates without significant contamination.

using the L_X-F_{XO} and the L_X-W_{12} selection, respectively; 2611 galaxies are selected by both methods. In total, we select 8667 galaxy candidates with a contamination fraction of $\sim 3\%$.

We identified 110 nearby galaxies with unidentified off-center X-ray sources in addition to the central source. We exclude them from the final catalog (see Appendix E for detailed descriptions of local galaxies). Additionally, we exclude 10 nearby ($z < 0.01$) optically faint ($L_r < 10^7 L_{r\odot}$) objects that are likely globular clusters.

6. The CSC2 X-Ray Galaxy Catalog

Our final CGC comprises 8547 galaxies: 24% are selected by optical spectroscopy, 30% by two methods using three photometric quantities (L_X , F_{XO} , and W_{12}), and 46% by one method using either (L_X-F_{XO}) or (L_X-W_{12}). The fractions of contaminations (primarily by QSOs) are $\sim 3\%$ and $\sim 5\%$ for the selection by (L_X-F_{XO}) and (L_X-W_{12}), respectively. The redshift of the CGC galaxies ranges from 0.04 and 0.7 (counting at the 5th and 95th percentiles) with a mean value of 0.3 ± 0.2 . The optical luminosity varies from 10^{10} to $10^{11} L_{r\odot}$, and the X-ray luminosity ranges from 2×10^{40} to 2×10^{43} erg s^{-1} (counting at the 5th and 95th percentiles).

For completeness and convenience, we add the 2MASS near-IR data and GALEX UV data to CGC. The 2MASS and GALEX catalogs are described in Section 3, and the same crossmatch procedure as in Section 4 is applied to find the counterparts.

One galaxy (2CXO J143456.7+340627), classified as a G2 star in the SDSS spectroscopic class, was first rejected in the spec- z sample but was reselected by both L_X-F_{XO} and L_X-W_{12} methods. We checked the optical images and confirmed that it is a spiral galaxy.

The columns in the CGC are described below, and the online catalog is available on the dedicated website¹⁸ and in a .tar.gz package.

1. *CSCID* CSC v2 name, e.g., 2CXO_J001625.8-270704. Note that a blank space between 2CXO and J was replaced by “_”.
2. *SDSSID* SDSS objID.
3. *PSID* Pan-STARRS objID.
4. *LEGSID* Legacy South ls_id.
5. *WISEID* WISE designation.
6. *2MASSID* 2MASS psc/xsc designation.
7. *GALEXID* GALEX objID.
8. *R.A., decl.* J2000 coordinates of the CSC 2 sources.

9. *sep_X{S, P, L, W}* separations in arcseconds between the positions of X-ray and {SDSS, Pan-STARRS, Legacy, WISE} sources.
10. *spec* 1 if the galaxy is selected by the SDSS spectroscopic class, 0 otherwise.
11. *photo* 1 if the galaxy with SDSS photo- z is selected by L_X-F_{XO} , 0 otherwise (see Section 5.1.3).
12. *wph* 1 if the galaxy with SDSS photo- z is selected by L_X-W_{12} , 0 otherwise (see Section 5.2.2).
13. *PS* 1 if the galaxy with Pan-STARRS photo- z is selected by L_X-F_{XO} , 0 otherwise (see Section 5.3).
14. *wPS* 1 if the galaxy with Pan-STARRS photo- z is selected by L_X-W_{12} , 0 otherwise (see Section 5.3).
15. *LS* 1 if the galaxy with Legacy South photo- z is selected by L_X-F_{XO} , 0 otherwise (see Section 5.3).
16. *wLS* 1 if the galaxy with Legacy South photo- z is selected by L_X-W_{12} , 0 otherwise (see Section 5.3).
17. *z* redshift.
18. *z_error* redshift error.
19. *z_code* source of the redshift. *z_SDSSsp* for SDSS spec- z ; *z_SDSSph* for SDSS photo- z ; *z_PSph* for Pan-STARRS photo- z ; *z_LEGSph* for Legacy photo- z ; *z_LEGSsp* for Legacy spec- z ; 2MRS for 2MASS redshift survey.
20. *d* luminosity distance in megaparsecs, assuming $H_0 = 69.6$ and $\Omega_M = 0.286$.
21. *r* optical magnitude in the *r* band.
22. *r_code* source of the *r* mag. *r_SDSS* for SDSS; *rPSkron* for Pan-STARRS kron mag; *rPSap* for Pan-STARRS aperture mag; *r_LEGS* for Legacy South.
23. *g* optical magnitude in the *g* band.
24. *g_code* source of the *g* mag. Same as in *r_code*.
25. *k* 2MASS *k*-band magnitude.
26. *W1* WISE magnitude in the W1-band ($3.4 \mu\text{m}$).
27. *W1snr* W1 S/N.
28. *W2* WISE magnitude in the W2-band ($4.6 \mu\text{m}$).
29. *w2snr* W2 S/N.
30. *W3* WISE magnitude in the W3-band ($12 \mu\text{m}$).
31. *w3snr* W3 S/N.
32. *W12* W1–W2 color.
33. *W23* W2–W3 color.
34. *nuv_mag* GALEX magnitude in the NUV band ($\lambda_{\text{eff}} = 2310 \text{ \AA}$).
35. *fuv_mag* GALEX magnitude in the FUV band ($\lambda_{\text{eff}} = 1528 \text{ \AA}$).
36. *significance* source flux significance from the CSC2 master source table.
37. *likelihood_class* True or Marginal from the CSC2 master source table.
38. *FxB15* X-ray flux in *B* band (0.5–7 keV) in unit of 10^{-15} erg s^{-1} cm^{-2} .
39. *FxS15* X-ray flux in *S* band (0.5–2 keV) in units of 10^{-15} erg s^{-1} cm^{-2} . This is the sum of CSC soft (0.5–1.2 keV) and medium (1.2–2 keV) band fluxes.
40. *FxH15* X-ray flux in *H* band (2–7 keV) in unit of 10^{-15} erg s^{-1} cm^{-2} .
41. *FxB15_lolim* lower limit of FxB15.
42. *FxB15_hilim* upper limit of FxB15.
43. *FxB_code* source of the *B*-band flux from CSC2. *flux_aper_b* = aperture flux in *B* band (95%). *flux_aper_a* = sum of the *S* and *H* fluxes if *flux_aper_b*

¹⁸ <https://cxc.cfa.harvard.edu/GalaxyAtlas/CGC/>

is unavailable (5%). flux_aper_w = aperture flux in *W* band for HRC only (0.1%).

44. *FxS_code* source of the *S*-band flux from CSC2.
45. *FxH_code* source of the *H*-band flux from CSC2.
46. *LxB40* X-ray luminosity in *B* band (0.5–7keV) in units of 10^{40} erg s⁻¹.
47. *LxS40* X-ray luminosity in *S* band (0.5–2keV) in units of 10^{40} erg s⁻¹.
48. *LxH40* X-ray luminosity in *H* band (2–7keV) in unist of 10^{40} erg s⁻¹.
49. *Lr10* Optical luminosity in *r* band in solar units, assuming $M_{r_{\odot}} = 4.68$ mag.
50. *Lk10* NIR luminosity in *k* band in solar units, assuming $M_{k_{\odot}} = 3.28$ mag.
51. *logFxo* $\log(Fx/Fo)$ (see Section 5.1.1).
52. *class_SDSS* SDSS spectroscopic class.
53. *subclass_SDSS* SDSS spectroscopic subclass.
54. *class_PS* Pan-STARRS class (galaxy, QSO, star).
55. *type_PS* Pan-STARRS type (*E* for early-type galaxies, *S* for spiral galaxies).
56. *type_LEG* Legacy type (Dev for the de Vaucouleurs profile; EXP for the exponential profile with a variable axis ratio; REX for the round exponential).
57. *p_cs* fraction of Galaxy Zoo votes for combined spiral (clockwise + counterclockwise + edge-on).
58. *p_el* fraction of Galaxy Zoo votes for elliptical.
59. *p_mg* fraction of Galaxy Zoo votes for merger.
60. *p_dk* fraction of Galaxy Zoo votes for do not know.

7. Discussion and Future Plans

The CGC provides a basis for the investigation of a number of critical scientific questions. We describe some examples below, including results from previous studies and investigations we plan to publish in the near future.

7.1. X-ray Scaling Relations in Early- and Late-type Galaxies

The X-ray emission of early- and late-type galaxies has different properties: (1) The X-ray emission from a gravitationally bound hot halo is the dominant source in typical giant elliptical galaxies, which is correlated with M_{TOTAL} (including dark matter), but not with M_* (or L_K). Additionally, low-mass X-ray binaries (LMXBs) whose total L_X is proportional to M_* or L_K are present in all early-type galaxies and become the dominant source in X-ray faint early-type galaxies (e.g., Boroson et al. 2011). (2) The hot ISM and high-mass X-ray binaries (HMXBs), directly related to star formation episodes, dominate in spiral galaxies (e.g., Lehmer et al. 2010; Li & Wang 2013). (3) AGNs (mostly low-luminosity AGNs—LLAGNs) may be present in some galaxies of both types.

In the CGC, we provide available information to separate the CGC galaxies into early and late types. From the SDSS spectroscopy database, we include the subclass—STARBURST, STARFROMING, or unclassified. The unclassified galaxies are likely to be early-type galaxies with no strong emission lines. We also include in the CGC the morphological types based on the citizen votes from the Galaxy Zoo catalog and those from the Pan-STARRS catalog of broad morphology by Goddard & Shamir (2020). We include from the DESI Legacy catalog the radial profile types—DEV (de Vaucouleurs profile), EXP (exponential profile with a variable axis ratio), or REX (round exponential). In future work, we plan to use these

morphology types and investigate the scaling relations and their implications separately in early- and late-type galaxies.

7.2. X-Ray Luminosity Function

The X-ray luminosity function (XLF) is a critical measurement that can be used to test theoretical models and numerical simulations. However, XLFs can only be built from large unbiased samples. Most XLFs published so far are based on a small number of (preferentially X-ray bright) galaxies. For example, Ptak et al. (2007) used 40 early-type and 46 late-type galaxies from the GOODS fields, and Tzanavaris & Georgantopoulos (2008) used 101 early-type and 106 late-type galaxies from the combined data of the Chandra deep fields and XBootes fields. These XLFs are highly unconstrained both at the low- and high-luminosity end and cannot convincingly constrain dependences on mass or redshift.

Using the CGC, we can apply the modified 1/VMAX (nonparametric) method and the (parametric) maximum likelihood techniques, with special attention to the Eddington bias and the uncertainty in photo-*z*, e.g., by using deconvolution- or convolution-based estimators. We can compare the observed XLF with cosmological simulations, which so far has only been done for the XRB XLFs (e.g., Fragos et al. 2013).

7.3. Rare Objects—XBONGs

The existence of an unusual population of XBONGs has been known since the Einstein mission (Elvis et al. 1981). Their X-ray emission is as luminous as that of typical AGNs ($L_X > 10^{42}$ erg s⁻¹), but they show no signs of AGN activity in their optical spectra. XBONGs are attracting more attention on both theoretical and observational grounds since Chandra and XMM-Newton observations have revealed a significant number of XBONG candidates (Fiore et al. 2000; Comastri et al. 2002; Georgantopoulos & Georgakakis 2005; Civano et al. 2016). However, the nature of XBONGs is yet to be understood.

One possibility is that they could be intrinsically luminous but heavily obscured AGNs, where the covering factor of the obscuration is significant enough that neither broad nor narrow lines escape. This scenario is interesting because they may be part of a missing population of hard X-ray sources necessary to explain the observed X-ray background emission (Gilli et al. 2007; Ueda et al. 2014). Another possible explanation is the dilution of nuclear emission lines by the bright starlight of the host galaxy (Moran et al. 2002), i.e., type 2 AGNs with stellar light bright enough to outshine the AGN signature. These galaxies are often called optically dull AGNs (OD AGNs).

Alternatively, XBONGs could be groups of galaxies with a large amount of intragroup medium (IGM) compared to the typical ISM of a single galaxy. These groups may not be recognized in typical optical observations if they are poor or even fossil groups, the end products of galaxy mergers dominated by a single elliptical galaxy (e.g., Ponman et al. 1994; Jones et al. 2003; also called an X-ray overluminous elliptical galaxy, or OLEG, by Vikhlinin et al. 1999 and isolated OLEG by Yoshioka et al. 2004).

Using the large CGC, we can explore XBONG candidates. The X-ray spectral shape (either by hardness ratios or absorbing N_H columns) is one of the critical quantities that can separate the obscured AGNs (hard) and hot gas (soft). The spatial extents can further help to identify hot gas and the temporal variations to identify AGNs. Additionally, WISE IR

colors can help to assess the degree of obscuration. Investigating the X-ray spectral, temporal, and spatial characteristics and WISE colors of the XBONG candidates, we will address the origin and nature of the unknown population (Kim et al. 2023).

Appendix A NWAY False-match Rate

We have run extensive simulations to determine the rate of false matches and to optimize the match statistics. The simulation is performed in the same way as in the real match (Section 4), ensemble by ensemble, but after shifting the source positions in eight directions (horizontal, vertical, and diagonal) by $30''$. Figure 2 shows the matched sources from each catalog in Galactic coordinates in the Aitoff projection. The number of total matches ranges from 40,000 to 110,000. The source density varies considerably across the sky.

Since the rate of false positives is a strong function of source density, we performed the simulations across a wide range of X-ray and optical and IR source densities. Figure A1 illustrates the distribution of the total number and density of sources in each ensemble from the catalogs. Note that CSC2 consists of 4380 ensembles, each covering a small part of the contiguous sky. The number of sources per ensemble ranges from 1 to $\sim 500,000$, and the source density extends from 0.0003 to ~ 300 sources per arcmin^2 . The source number and density of the data sets used in the simulations are also marked (color coded) at the top of the two panels of Figure A1.

Figure A2 shows the estimated false-positive rates as a function of optical and IR source density.

The rate of false positives is determined by dividing the average number of matches found with the simulation for a given ensemble (fake matches) by the number of matches identified with the original source positions (real matches). Because the X-ray source density is always low (<3 in arcmin^2), the false-match rate primarily depends on the optical and IR source density. The top left panel of Figure A2 indicates the false-positive rate when no additional selection in

separation is applied, i.e., with all counterparts with separation $<10''$ set by the initial parameter. When the source density is low (<10 sources in arcmin^2), the rate of false positives is also low ($<10\%$). When the source density is intermediate (10–30 sources in arcmin^2), the rate of false positives is about $\sim 20\%$. When the source density is very high (>40 sources in arcmin^2), the rate of false positives is also very high ($\sim 50\%$).

The top right panel of Figure A2 shows the rates after applying a separation $<3''$. The false-positive rate is significantly reduced (almost by half). The rate is always lower than 5% when the source density is lower than 10. The horizontal line in the figure indicates the 5% level. When the source density is intermediate (10–30), the rate of false positives is about $\sim 10\%$. When the source density is very high (>40), the false rate is still as high as 40%. Based on the simulation results, we choose (1) to apply the final cut at separation $<3''$ and (2) to exclude the regions of very high source densities (see below).

For the second purpose, we remove the Galactic plane ($|b| > 15$), entire ensembles with high optical or IR source density, such as the nearby large galaxies, M31 and M33, and parts of ensembles within a given radius from known objects, such as Galactic globular clusters (e.g., M2, M3, and M5). In our matches between CSC2 and the optical and IR catalogs, we achieve a rate of the false-positive match of $\sim 5\%$ after excluding these crowded fields and applying our strict selection criteria.

To select the optimal NWAY parameter, p_{any} (the probability that one of the associations is correct), we test how the false-match rate decreases with increasing p_{any} . Increasing p_{any} from $p_{\text{any}} > 0.5$ to $p_{\text{any}} > 0.8$ (the bottom panel of Figure A2), we find that the number of false matches decreases, but the number of real matches decreases as well, in a way that the reduction of the false-positive rate is only minimal. Therefore, we choose $p_{\text{any}} > 0.5$ to select the final matches.

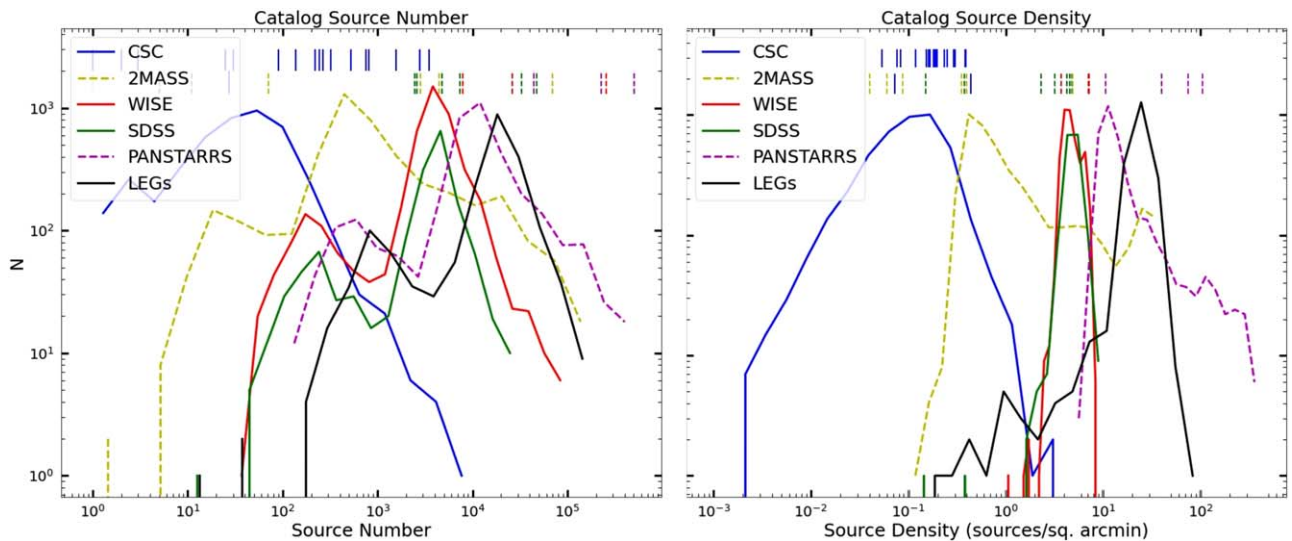


Figure A1. The total number and density (in arcmin^2) of X-ray, optical, and IR sources in each ensemble. Unique colors and line styles mark different catalogs (see legend). The ensembles used in the simulation are marked by the color and line style of the catalogs by the small vertical bars at the top of the figure.

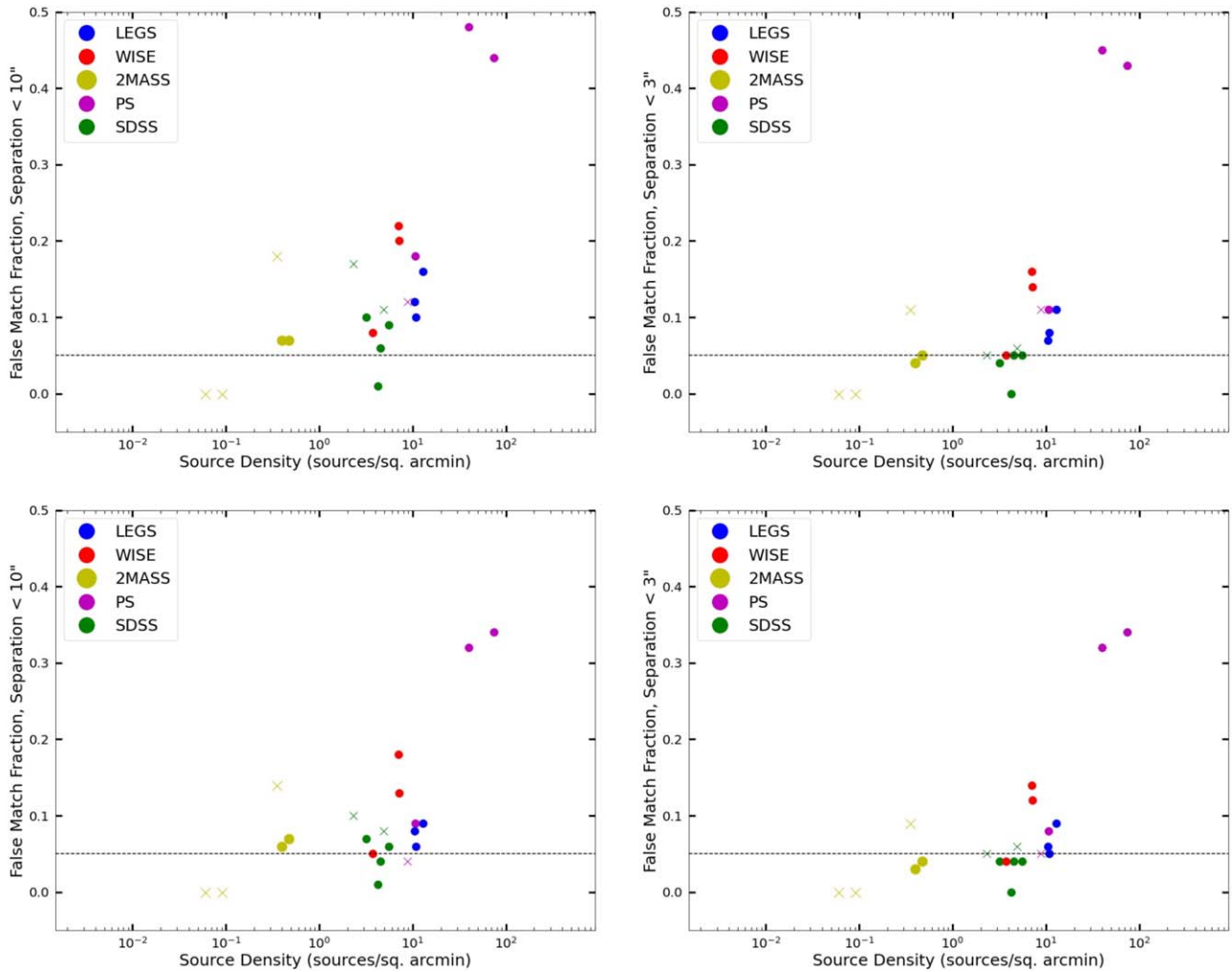


Figure A2. False-match fractions as a function of optical and IR source density. (Top left) Applying the initial maximum separation of $10''$ without further selection by separation, and (top right) applying an additional selection with a separation $< 3''$. In both cases, we set $p_{\text{any}} > 0.5$. (Bottom panel) Same as the top panel, but with $p_{\text{any}} > 0.8$. The crosses indicate the ensembles with a relatively small number of counterparts ($\lesssim 10$), which are less significant. The horizontal lines indicate that the false-positive rate is 5%, a typical limit in this work.

Appendix B Galaxy Zoo Morphological Types

We extracted morphological types from the Galaxy Zoo 1 release in the SDSS Zoo-Votes table (Lintott et al. 2008). In the spec- z and photo- z samples, we find 1249 objects with listed morphological types. The Zoo-Votes table contains the fraction of votes for six categories: ellipticals, spirals (clockwise, counterclockwise, and edge-on), mergers, and stars/do not know. We classified objects with a fraction of votes > 0.5 as belonging to that category. We note that the Zoo classifications with vote fractions below 0.7 may be less reliable. The number of sources that fall into each morphological type is listed in Table B1. About three-quarters are elliptical galaxies. This X-ray and Zoo-selected sample is somewhat biased against less massive spiral galaxies.

Table B1
SDSS Morphological Types

| Type | No. |
|--|-----|
| Ellipticals ($p_{\text{el}} > 0.5$) | 805 |
| Spirals ($p_{\text{cs}} > 0.5$) | 226 |
| Mergers ($p_{\text{mg}} > 0.5$) | 38 |
| Do not knows ($p_{\text{dk}} > 0.5$) | 17 |

Appendix C Comparison of Redshifts

The spectroscopic redshifts (spec- z) and photometric redshifts (photo- z) for the SDSS counterparts of CSC2 sources are taken from the SDSS DR16 SpecObj and Photo- z tables: 2959

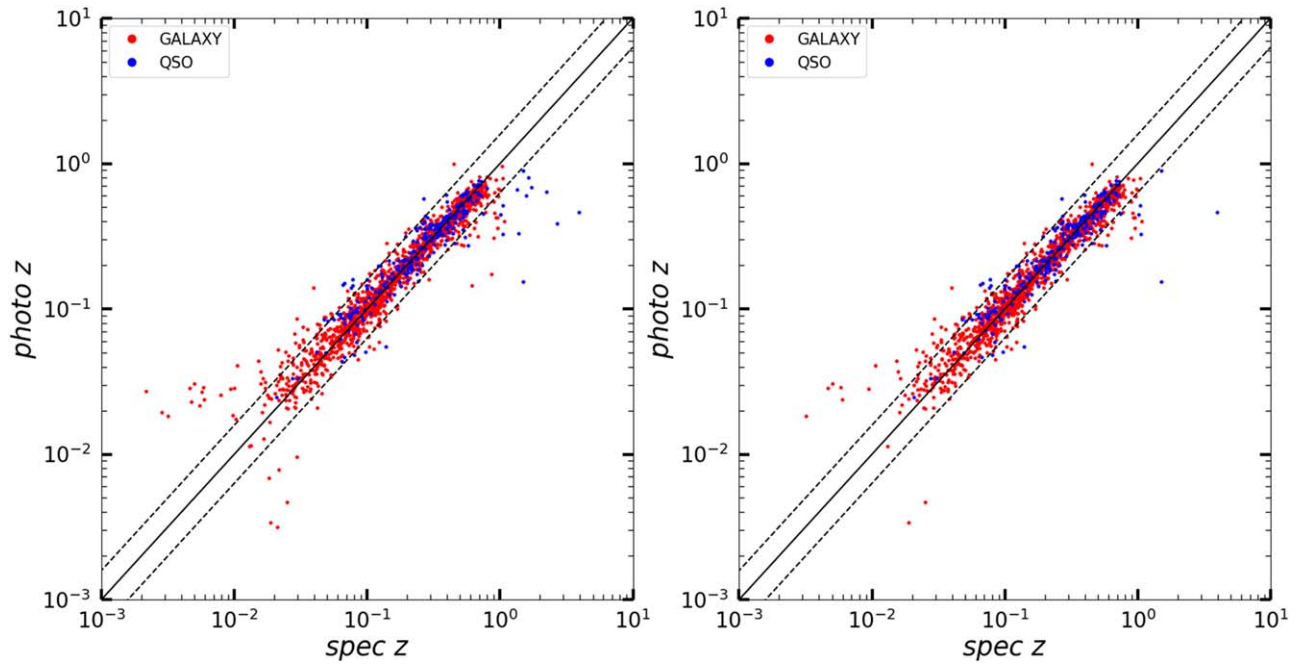


Figure C1. (Left) Comparison of SDSS spec- z and photo- z . (Right) Poor-quality measurements are excluded by selecting only those with $zWarning = 0$ (for spec- z), $photoErrorClass = 1$ (for photo- z), and $z / error(z) > 2$ (for all).

Table C1

Fraction of Outliers when Compared with SDSS Spec- z

| | SDSS Photo- z | | Pan-STARRS Photo- z | | Legacy Photo- z | |
|--|-----------------|-----------|-----------------------|-----------|-------------------|----------|
| | Total | Outlier | Total | Outlier | Total | Outlier |
| | Galaxies | 2141 | 144 (7%) | 1649 | 176 (11%) | 917 |
| QSOs | 813 | 171 (21%) | 339 | 104 (31%) | 286 | 48 (17%) |
| Stars | 5 | 5 (100%) | 3 | 3 (100%) | 2 | 2 (100%) |
| After screening for quality control ^a | | | | | | |
| Galaxies | 1926 | 87 (5%) | 1550 | 134 (9%) | 915 | 42 (5%) |
| QSOs | 619 | 67 (11%) | 280 | 67 (24%) | 282 | 47 (17%) |
| Stars | 2 | 2 (100%) | 0 | 0 | 1 | 1 (100%) |

Notes.

^a Screening for quality control, $zWarning = 0$ for SDSS spec- z , $photoErrorClass = 1$ for SDSS photo- z , $extrapolationClass = 0$ and $extrapolationPhotoz = 0$ for Pan-STARRS photo- z .

objects have both spec- z and photo- z available. In Figure C1 photo- z is plotted against spec- z . The dashed lines indicate the difference of 0.2 dex, i.e., $|\log(photo-z) - \log(spec-z)| = 0.2$. Different types are color coded. Overall, photo- z follows spec- z well in the mid-range of $z = 0.03-0.5$. However, photo- z deviates from spec- z at lower z (< 0.03 ; primarily for galaxies), where photo- z is often greater than spec- z . At higher z (> 0.5 , primarily for QSOs), photo- z is often lower than spec- z . In Table C1 we list the total number of sources and the fraction of outliers beyond 0.2 dex. Compared with the SDSS spec- z of galaxies, 7% of the SDSS photo- z are inconsistent. This outlier fraction is considerably higher for QSOs (21%).

In the right panel of Figure C1, we select only those with high-quality measurements. First, we apply $zWarning = 0$ for spec- z and $photoErrorClass = 1$ for photo- z (see Beck et al. 2016). This excludes $\sim 15\%$ in this sample with both spec- z and photo- z . Among the objects with only photo- z , the fraction is significantly higher at $\sim 50\%$. Nonetheless, we apply this strict selection rule. Then, we apply $error(z)/z < 0.5$ to further

Table C2

Fraction of Outliers when Compared with SDSS Photo- z

| Pan-STARRS Total | Photo- z Outlier | Legacy Photo- z | |
|--|--------------------|-------------------|----------|
| | | Total | Outlier |
| 3217 | 178 (6%) | 2567 | 183 (7%) |
| After screening for quality control ^a | | | |
| 2980 | 141 (5%) | 2551 | 176 (7%) |

Note.

^a Screening for quality control, $zWarning = 0$ for SDSS spec- z , $photoErrorClass = 1$ for SDSS photo- z , $extrapolationClass = 0$ and $extrapolationPhotoz = 0$ for Pan-STARRS photo- z .

remove objects with significant uncertainties. This condition removes $\sim 1\%$ of the sample. As in Table C1, the outlier fraction is reduced to 5% for galaxies and 11% for QSOs. Note that this 5% is comparable to the similar fraction of the false-match rate already given when crossmatching CSC2 and other catalogs. While we try to reduce these uncertainties, we consider them an inevitable limit.

The photo- z data for Pan-STARRS and Legacy are from Beck et al. (2021) and Zou et al. (2019), respectively. These data are compared with SDSS spec- z in Figure C2. In addition to the screening described above, we only select those with no extrapolation in Pan-STARRS photo- z (by $extrapolationClass = 0$ and $extrapolationPhotoz = 0$; see Beck et al. 2021). The two photo- z data are consistent with the SDSS spec- z (as good as the SDSS photo- z). The galaxy outlier fractions are 9% and 5% for Pan-STARRS and Legacy, respectively. Again, the outlier fraction for QSOs is higher (17–24%).

We also compare the two photo- z data with SDSS photo- z in Figure C3. No information about their classes is available for these samples, only the expectation that they primarily consist of galaxies. The photo- z data are consistent with the SDSS photo- z . The outlier fractions are 5% and 7% for Pan-STARRS and Legacy, respectively (Table C2).

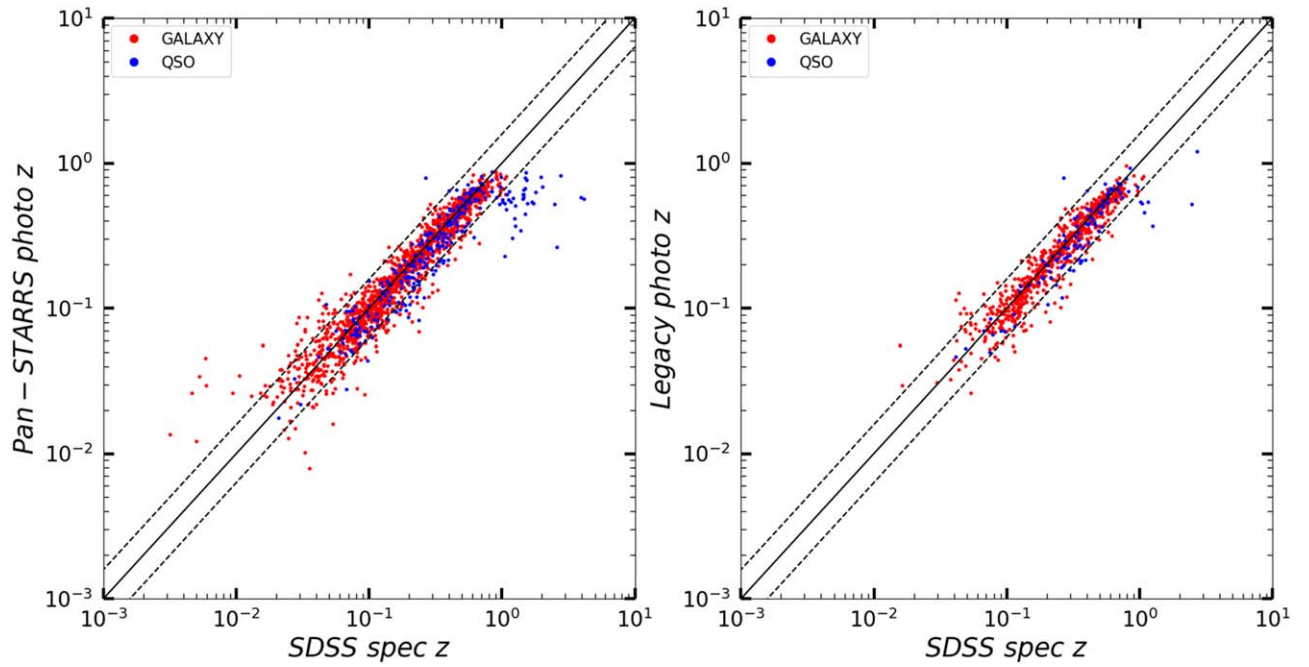


Figure C2. (Left) Pan-STARRS photo- z and (right) Legacy photo- z are compared with SDSS spec- z . Poor-quality measurements are excluded by selecting only those with $zWarning = 0$ (for SDSS spec- z), $extrapolationClass = extrapolationPhotoz = 0$ (for Pan-STARRS photo- z), and $error(z)/z < 0.5$ (for all).

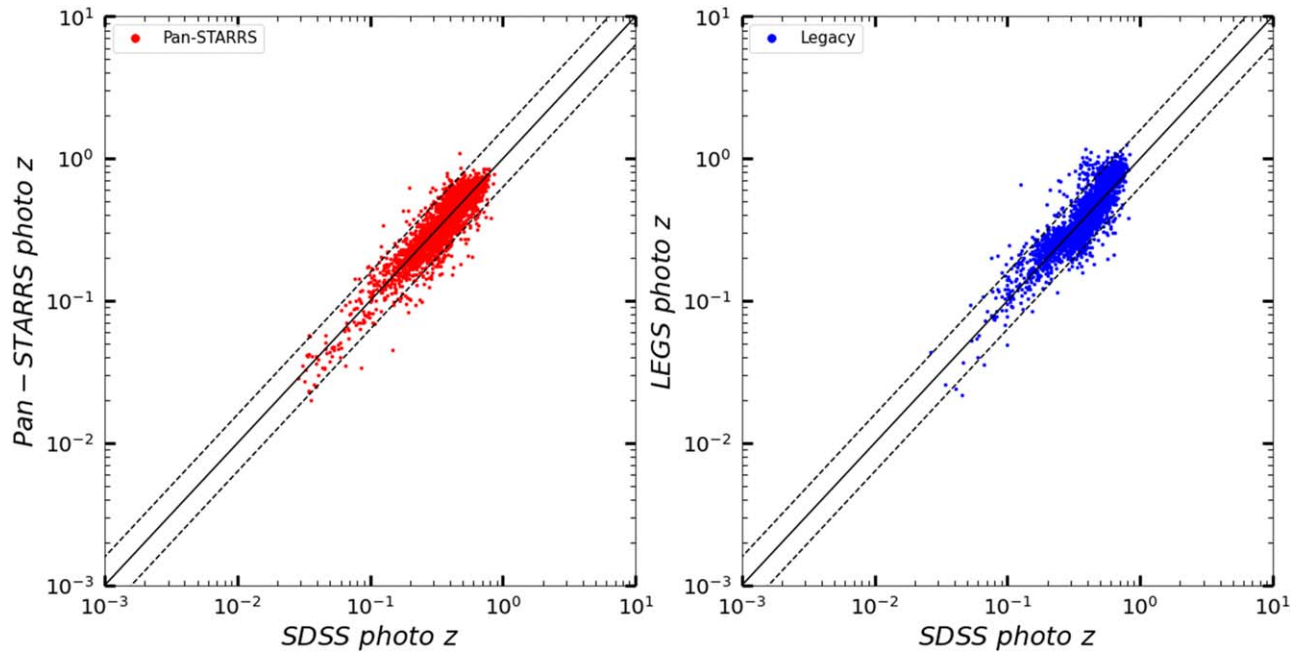


Figure C3. (Left) Pan-STARRS photo- z and (right) Legacy photo- z are compared with SDSS photo- z . Poor-quality measurements are excluded by selecting only those with $photoErrorClass = 1$ (for SDSS photo- z), $extrapolationClass = extrapolationPhotoz = 0$ (for Pan-STARRS photo- z), and $error(z)/z < 0.5$ (for all).

Appendix D Comparison of Optical Magnitudes

We compare the optical r mag from SDSS, Pan-STARRS, and Legacy catalogs in Figure D1. We take r from the SDSS PhotoObj table.¹⁹ This is the model magnitude to consider the

extended source emission beyond the PSF properly. Similarly, we take $rMeanKronMag$ from the Pan-STARRS MeanObject table (Flewelling et al. 2020) and mag_r from the Legacy $ls_dr8.tractor_s$ table.²⁰ The mean difference is 0.15 ± 0.4 (0.2 ± 0.2) between SDSS and Pan-STARRS (Legacy). Note that the scatter is small at the bright end ($r < 19$ mag), but

¹⁹ <https://skyserver.sdss.org/dr16/en/help/browser/browser.aspx#&&history=description+PhotoObjAll+U>

²⁰ <https://www.legacysurvey.org/dr8/description/>

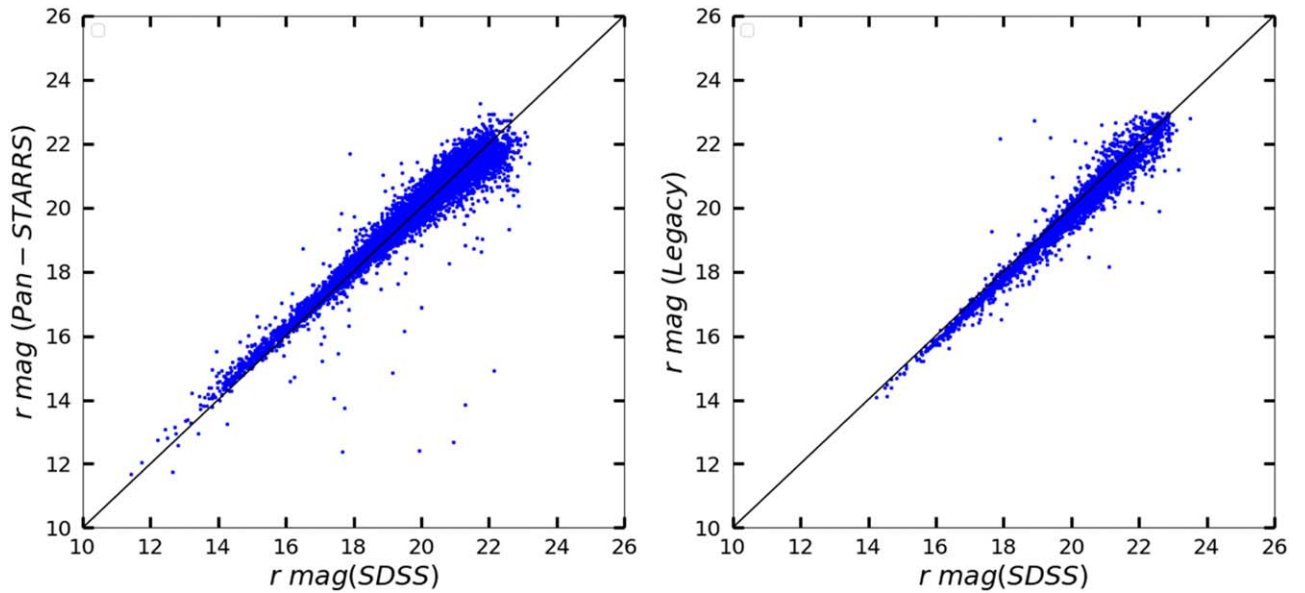


Figure D1. (Left) The Pan-STARRS r mag and (right) the Legacy r mag are compared with the SDSS r mag.

significantly increases at the faint end ($r > 19$ mag). Overall, 90% of them are consistent within 0.5 mag.

Appendix E Local Galaxy Exclusion

Our matching procedure assumes that each galaxy corresponds to one CSC2 source. If multiple CSC2 sources belong to one galaxy, our crossmatching procedure would have listed the central source and underestimated the total X-ray flux. Unlike distant galaxies in typical deep surveys (e.g., Chandra Deep Field or COSMOS), local galaxies, which may have multiple X-ray sources, need special attention.

First, we have excluded all the CSC2 sources with an extent $>30''$ (see Section 2), i.e., most nearby galaxies within ~ 100 Mpc or $z < 0.025$, where extended and complex X-ray emission is detected with Chandra. These nearby galaxies contain multiple X-ray sources, including diffuse hot-gas X-ray binaries (HMXBs and LMXBs), and possibly low-luminosity AGN. We refer to the Chandra Galaxy Atlas (Kim et al. 2019) and XMM-Newton Galaxy Atlas (Islam et al. 2021) for more information on the nearby galaxies.

Second, for the galaxies within a few times 100 Mpc (roughly $z < 0.1$), there may be additional off-center CSC2 X-ray sources that could be detected but are not included in the match, underestimating the total galaxy flux and luminosity. To estimate how many off-center X-ray sources may be missed in our galaxy sample, we have examined the 2MASS redshift survey (2MRS), which includes 44599 local ($z \lesssim 0.1$) galaxies (Huchra et al. 2012) with 98% completeness down to $K_s = 11.75$ mag.

We find that 417 2MRS galaxies have our selected galaxy candidates within their boundaries (the total magnitude extrapolation radius in 2MRS). In 128 of these galaxies, the CSC2 source lies at large radii $>3''$ from the center. These X-ray sources are likely background galaxies as identified in the spec- z or photo- z samples (see Section 5). In 160 galaxies, we find only one central CSC2 source, and in 129 galaxies, we find a central CSC2 source plus an additional off-center CSC2 source. For 19 of the latter, the off-center sources are identified

with background galaxies or QSOs in the SDSS spec- z sample (see Section 5.1). In the remaining 110 galaxies, the off-center sources do not have a crossmatch counterpart. Conservatively, we exclude these 110 galaxies from the CGC because the unidentified off-center sources may belong to the galaxies. We will present local galaxies with multiple X-ray sources in a separate paper.

Comparing the 2MRS 289 galaxies with a central CSC2 source with the SDSS values, we found that 136 of them have redshifts in agreement with SDSS spec- z , within 1%. One exception is IC 2475, where the difference is 30%, for which we adopted the SDSS spec- z value that is consistent with other measurements available in the NED. For the remaining 153 2MRS galaxies, there is no SDSS spec- z , and the photo- z values (including 131 from Pan-STARRS) are different from the 2MRS spec- z . Of these, 66 are included in the 110 galaxies that we have excluded from our catalog (see above). For the remaining 87 galaxies, we adopted the 2MRS redshifts.

We realized that some globular clusters (e.g., from the SLUGGS Survey; Forbes et al. 2017) in nearby galaxies are spectroscopically classified as galaxies in the SDSS SpecObj table. Because of their proximity and low optical luminosity, we exclude 10 nearby ($z < 0.01$) optically faint ($L_r < 10^7 L_{r\odot}$) objects. All of them are from the SDSS spec- z sample.

Appendix F Redshift and L_X Distributions of Galaxies from Different Samples

In this appendix, we test whether different samples (e.g., spec- z versus photo- z samples) have different global properties (e.g., in redshift and L_X) that in turn affect the galaxy-finding statistics (as in Tables 5, 8, 10, 11). In Figure F1 (left), we show the z distributions of different samples, including the spec- z sample from SDSS and the photo- z samples from SDSS, Pan-STARRS, and Legacy. The z distributions do not vary significantly from one sample to the next, particularly at higher z ($z > 0.3$). The only noticeable difference is a slightly higher fraction of galaxies at lower z ($z < 0.1$) in the SDSS spec- z and

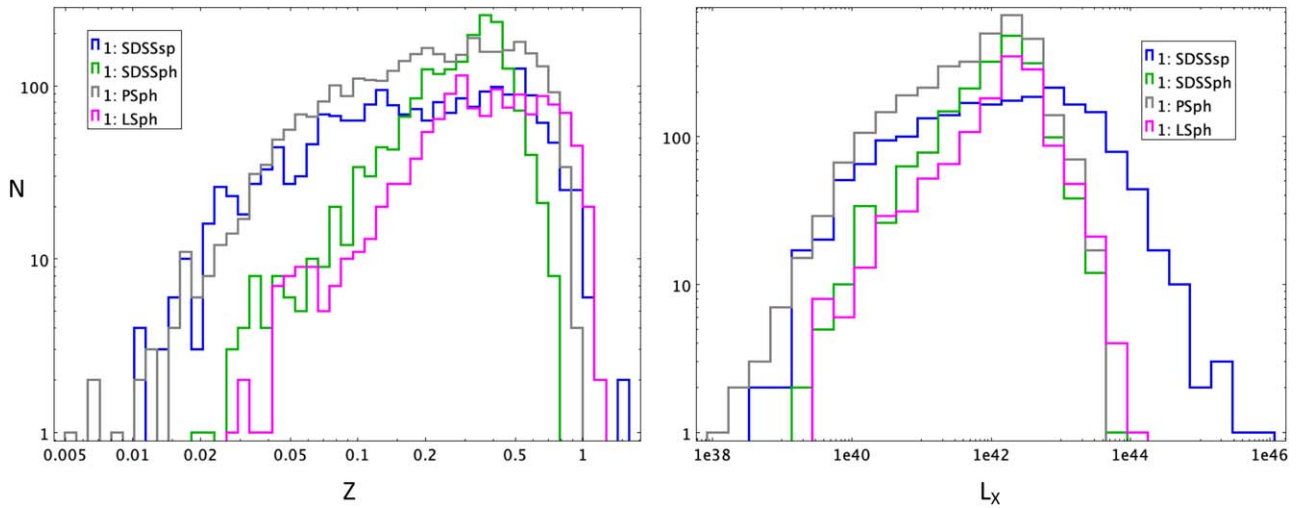


Figure F1. (Left) The z distributions and (right) the $\log(L_X)$ distributions of the spec- z sample (SDSS) and the photo- z samples (SDSS, Pan-STARRS, and Legacy).

Table F1
The z and $\log(L_X)$ Distributions of Different Samples

| | z | | $\log(L_X)$ | |
|-----------------|------|--------|-------------|-------|
| | Mean | std | Mean | std |
| All | 0.31 | (0.21) | 41.9 | (0.9) |
| SDSS spec- z | 0.27 | (0.23) | 42.1 | (1.2) |
| SDSS photo- z | 0.31 | (0.13) | 42.0 | (0.7) |
| PS photo- z | 0.28 | (0.20) | 41.8 | (0.8) |
| LS photo- z | 0.43 | (0.24) | 42.1 | (0.7) |

Pan-STARRS samples than in the SDSS photo- z and Legacy-S samples. In Table F1 we list the mean and the standard deviation (std) of the z distributions. The mean values are similar within half of a std. We also applied a few different z ranges to examine whether our galaxy selection statistics depend on the different z distributions between the spec- z and photo- z samples. As expected, based on the insignificant difference among different samples, we do not find significantly different statistics.

Similarly, in Figure F1 (right), we show the $\log(L_X)$ distributions of different samples. Again, they are similar, with one noticeable exception—all galaxies with higher L_X ($>10^{44}$ erg s $^{-1}$) are from the SDSS spec- z sample. These high L_X galaxies could not be selected in the photo- z samples because of the significant QSO contaminations. These are the XBONGs discussed in Kim et al. (2023). The mean and std of the $\log(L_X)$ distributions of different samples are listed in Table F1. Again, the mean values are similar within one-third of a std.

ORCID iDs

Dong-Woo Kim <https://orcid.org/0000-0002-7386-944X>

Giuseppina Fabbiano <https://orcid.org/0000-0002-3554-3318>

Juan Rafael Martinez Galarza <https://orcid.org/0000-0002-5069-0324>

Ewan O'Sullivan <https://orcid.org/0000-0002-5671-6900>

Arnold Rots <https://orcid.org/0000-0003-2377-2356>

References

Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, *ApJ*, 772, 26
Beck, R., Dobos, L., Buda'ari, T., et al. 2016, *MNRAS*, 460, 1371

- Beck, R., Szapudi, I., Flewelling, H., et al. 2021, *MNRAS*, 500, 1633
Bianchi, L., Shiao, B., & Thilker, D. 2017, *ApJS*, 230, 24
Bogdan, A., & Goulding, A. D. 2015, *ApJ*, 800, 124
Bolton, A. S., Schlegel, D. J., Aubourg, E., et al. 2012, *AJ*, 144, 144
Borson, B., Kim, D.-W., & Fabbiano, G. 2011, *ApJ*, 729, 12
Choi, E., Ostriker, J. P., Naab, T., et al. 2017, *ApJ*, 844, 31
Ciotti, L., Pellegrini, S., Negri, A., et al. 2017, *ApJ*, 835, 15
Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, 819, 62
Comastri, A., Mignoli, M., Ciliegi, P., et al. 2002, *ApJ*, 571, 771
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168
Elvis, M., Schreier, E. J., Tonry, J., Davis, M., & Huchra, J. P. 1981, *ApJ*, 246, 20
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, *ApJS*, 189, 37
Fabbiano, G., & Elvis, M. 2022, in *Handbook of X-ray and Gamma-ray Astrophysics*, ed. C Bambi & A Santangelo (Berlin: Springer)
Fabbiano, G. 2019, in *The Chandra X-ray Observatory*, ed. B. Wilkes & W. Tucker (Bristol: IOP Publishing), 7-1
Fabian, A. C. 2012, *ARA&A*, 50, 455
Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, *ApJS*, 251, 7
Fiore, F., La Franca, F., Vignali, C., et al. 2000, *NewA*, 5, 143
Forbes, D. A., Alabi, A., Brodie, J. P., et al. 2017, *AJ*, 153, 114
Fragos, T., Lehmer, B., Tremmel, M., et al. 2013, *ApJ*, 764, 41
Georgantopoulos, I., & Georgakakis, A. 2005, *MNRAS*, 358, 131
Gilli, R., Comastri, A., & Hasinger, G. 2007, *A&A*, 463, 79
Goddard, H., & Shamir, L. 2020, *ApJS*, 251, 28
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *ApJS*, 199, 26
Islam, N., Kim, D.-W., Lin, K., et al. 2021, *ApJS*, 256, 22
Jarrett, T. H., Cluver, M. E., Magoulas, C., et al. 2017, *ApJ*, 836, 182
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, *ApJ*, 735, 112
Jones, L. R., Ponman, T. J., Horton, A., et al. 2003, *MNRAS*, 343, 627
Kelly, A. J., Jenkins, A., & Frenk, C. S. 2021, *MNRAS*, 502, 2934
Kelly, B. C. 2007, *ApJ*, 665, 1489
Kim, D.-W., Anderson, C., Burke, D., et al. 2019, *ApJS*, 241, 36
Kim, D.-W., Barkhouse, W. A., Romero-Colmenero, E., et al. 2006, *ApJ*, 644, 829
Kim, D.-W., & Fabbiano, G. 2010, *ApJ*, 721, 1523
Kim, D.-W., Fabbiano, G., Brassington, N. J., et al. 2009, *ApJ*, 703, 829
Kim, D.-W., Fabbiano, G., & Trinchieri, G. 1992, *ApJ*, 393, 134
Kim, D.-W., Malnati, A., Cassity, A., et al. 2023, *ApJ*, submitted
Kim, D.-W., & Pellegrini, S. 2012, *Hot Interstellar Matter in Elliptical Galaxies* (Berlin: Springer)
Kim, M., Kim, D.-W., Wilkes, B. J., et al. 2007, *ApJS*, 169, 401
Kron, R. G. 1980, *ApJS*, 43, 305
Lehmer, B. D., Alexander, D. M., Bauer, F. E., et al. 2010, *ApJ*, 724, 559
Lehmer, B. D., Basu-Zych, A. R., Mieno, S., et al. 2016, *ApJ*, 825, 7
Li, J.-T., & Wang, Q. D. 2013, *MNRAS*, 435, 3071
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, 389, 1179
Maccararo, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, *ApJ*, 326, 680
Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, 817, 34
Mineo, S., Gilfanov, M., & Sunyaev, R. 2012, *MNRAS*, 419, 2095

- Moran, E. C., Filippenko, A. V., & Chornock, R. 2002, *ApJ*, 579, L71
- Moran, E. C., Lehnert, M. D., & Helfand, D. J. 1999, *ApJ*, 526, 649
- Nardini, E., Kim, D.-W., & Pellegrini, S. 2022, in *Handbook of X-ray and Gamma-ray Astrophysics*, ed. C. Bambi & A. Santangelo (Berlin: Springer)
- Ponman, T. J., Allan, D. J., Jones, L. R., et al. 1994, *Nature*, 369, 462
- Ptak, A., Mobasher, B., Hornschemeier, A., et al. 2007, *ApJ*, 667, 826
- Salvato, M., Buchner, J., Budavári, T., et al. 2018, *MNRAS*, 473, 4937
- Shapley, A., Fabbiano, G., & Eskridge, P. B. 2001, *ApJS*, 137, 139
- Stern, D., Assef, R. J., Benford, D. J., et al. 2012, *ApJ*, 753, 30
- Tananbaum, H., Avni, Y., Branduardi, G., et al. 1979, *ApJ*, 234, L9
- Tzanavaris, P., & Georgantopoulos, I. 2008, *A&A*, 480, 663
- Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, *ApJ*, 786, 104
- Vikhlinin, A., McNamara, B. R., Hornstrup, A., et al. 1999, *ApJ*, 520, L1
- Vogelsberger, M., Marinacci, F., Torrey, P., et al. 2020, *NatRP*, 2, 42
- Wilkes, B., & Tucker, W. 2019, *The Chandra X-ray Observatory: Exploring the High Energy Universe* (Bristol: IOP Publishing)
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, K., et al. 2010, *AJ*, 140, 1868
- Yoshioka, T., Furuzawa, A., Takahashi, S., et al. 2004, *AdSpR*, 34, 2525
- Zombeck, M. V. 1990, *Handbook of Space Astronomy and Astrophysics* (Cambridge: Cambridge Univ. Press)
- Zou, H., Gao, J., Zhou, X., et al. 2019, *ApJS*, 242, 8