


















# The eROSITA Final Equatorial-Depth Survey (eFEDS)

## Identification and characterization of the counterparts to point-like sources<sup>★</sup>

M. Salvato<sup>1,2</sup>, J. Wolf<sup>1,2</sup>, T. Dwelly<sup>1</sup>, A. Georgakakis<sup>3</sup>, M. Brusa<sup>4,5</sup>, A. Merloni<sup>1</sup>, T. Liu<sup>1</sup>, Y. Toba<sup>6,7,8</sup>, K. Nandra<sup>1</sup>, G. Lamer<sup>9</sup>, J. Buchner<sup>1</sup>, C. Schneider<sup>10</sup>, S. Freund<sup>10</sup>, A. Rau<sup>1</sup>, A. Schwobe<sup>9</sup>, A. Nishizawa<sup>11</sup>, M. Klein<sup>12</sup>, R. Arcodia<sup>1</sup>, J. Comparat<sup>1</sup>, B. Musiimenta<sup>4,5</sup>, T. Nagao<sup>8</sup>, H. Brunner<sup>1</sup>, A. Malyali<sup>1</sup>, A. Finoguenov<sup>1</sup>, S. Anderson<sup>13</sup>, Y. Shen<sup>14</sup>, H. Ibarra-Medel<sup>14</sup>, J. Trump<sup>15</sup>, W. N. Brandt<sup>16,17,18</sup>, C. M. Urry<sup>19</sup>, C. Rivera<sup>9</sup>, M. Krumpe<sup>9</sup>, T. Urrutia<sup>9</sup>, T. Miyaji<sup>20</sup>, K. Ichikawa<sup>1,21,22</sup>, D. P. Schneider<sup>16,17</sup>, A. Fresco<sup>1</sup>, T. Boller<sup>1</sup>, J. Haase<sup>1</sup>, J. Brownstein<sup>23</sup>, R. R. Lane<sup>24</sup>, D. Bizyaev<sup>25</sup>, and C. Nitschelm<sup>26</sup>

(Affiliations can be found after the references)

Received 25 June 2021 / Accepted 7 March 2022

### ABSTRACT

**Context.** In November 2019, eROSITA on board of the Spektrum-Roentgen-Gamma (SRG) observatory started to map the entire sky in X-rays. After the four-year survey program, it will reach a flux limit that is about 25 times deeper than ROSAT. During the SRG performance verification phase, eROSITA observed a contiguous 140 deg<sup>2</sup> area of the sky down to the final depth of the eROSITA all-sky survey (eROSITA Final Equatorial-Depth Survey; eFEDS), with the goal of obtaining a census of the X-ray emitting populations (stars, compact objects, galaxies, clusters of galaxies, and active galactic nuclei) that will be discovered over the entire sky.

**Aims.** This paper presents the identification of the counterparts to the point sources detected in eFEDS in the main and hard samples and their multi-wavelength properties, including redshift.

**Methods.** To identify the counterparts, we combined the results from two independent methods (NWAY and ASTROMATCH), trained on the multi-wavelength properties of a sample of 23k *XMM-Newton* sources detected in the DESI Legacy Imaging Survey DR8. Then spectroscopic redshifts and photometry from ancillary surveys were collated to compute photometric redshifts.

**Results.** Of the eFEDS sources, 24 774 of 27 369 have reliable counterparts (90.5%) in the main sample and 231 of 246 sources (93.9%) have counterparts in the hard sample, including 2514 (3) sources for which a second counterpart is equally likely. By means of reliable spectra, *Gaia* parallaxes, and/or multi-wavelength properties, we have classified the reliable counterparts in both samples into Galactic (2695) and extragalactic sources (22 079). For about 340 of the extragalactic sources, we cannot rule out the possibility that they are unresolved clusters or belong to clusters. Inspection of the distributions of the X-ray sources in various optical/IR colour-magnitude spaces reveal a rich variety of diverse classes of objects. The photometric redshifts are most reliable within the KiDS/VIKING area, where deep near-infrared data are also available.

**Conclusions.** This paper accompanies the eROSITA early data release of all the observations performed during the performance and verification phase. Together with the catalogues of primary and secondary counterparts to the main and hard samples of the eFEDS survey, this paper releases their multi-wavelength properties and redshifts.

**Key words.** methods: data analysis – X-rays: general – catalogues – surveys – galaxies: active – galaxies: distances and redshifts

## 1. Introduction

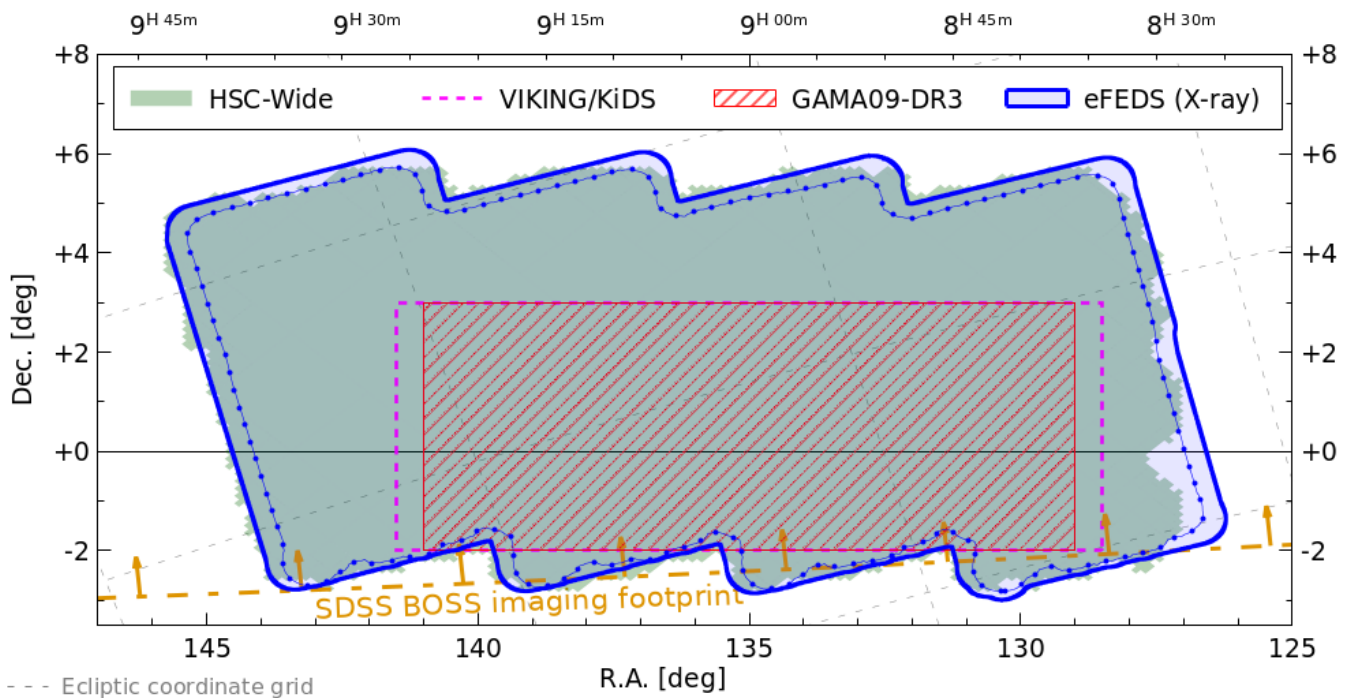
Across the electromagnetic spectrum, sensitive wide-area surveys serve multiple purposes. First and foremost, they help astronomers to draw a map of our cosmic neighbourhood. In doing so, they reveal the inner workings of the Milky Way, the local group, and the filamentary large-scale structure underpinning the distribution of matter. Secondly, by observing and cataloguing large numbers of stars, galaxies, groups, clusters, and superclusters of galaxies that are the main visible tracers of this large-scale structure, wide area surveys also provide new statistical tools for the study of classes and populations of astronomical objects, thus helping astronomers to better understand their life cycles, interactions, and ultimately, their physical properties.

X-ray surveys in particular reveal fundamental physical processes that are invisible at other wavelengths. Examples are the hot, diffuse plasma that virialises and thermalises within massive dark matter knots; accretion of matter onto compact objects, both Galactic and extragalactic; and the magnetic coronae of mostly young, fast-rotating stars. These are all phenomena that are accessible by X-ray sensitive instruments.

extended ROentgen Survey with an Imaging Telescope Array (eROSITA, Predehl et al. 2021) on board the Spektrum-Röntgen-Gamma (SRG) mission (Sunyaev et al. 2021), was designed to provide sensitive X-ray imaging and spectroscopy over a large field of view, thus unlocking unprecedented capabilities for surveying large areas of the sky to deep flux levels. Moreover, the SRG mission plan includes a long (four years), uninterrupted all-sky survey program (the eROSITA All-Sky Survey: eRASS; Predehl et al. 2021) capable of detecting millions of X-ray sources for the first time.

In order to demonstrate these ground-breaking survey capabilities and prepare for the science exploitation of the

<sup>★</sup> The data are only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](https://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/661/A3>



**Fig. 1.** eFEDS X-ray and multi-wavelength coverage. The thick blue line shows the outer bound of the region that was searched for X-ray sources. The thin beaded blue line shows the region with at least 500 seconds of effective X-ray exposure depth. We indicate the approximate coverage of several selected surveys that are particularly important for this work: Subaru HSC-Wide (shaded green region), KiDS/VIKING (dashed magenta box), GAMA09-DR3 (hatched red box). The eFEDS field is also covered in several other important surveys that completely (or almost completely) enclose the displayed region: e.g., the *Galex* all-sky surveys (in the UV), *Gaia* (in optical), Legacy Survey DR8 (optical combined with *Gaia* and WISE), VHS, and UKIDSS (in the near-infrared), WISE/NEOWISE-R (in the mid-infrared), and the SDSS (optical imaging and spectroscopy).

upcoming all-sky survey, the contiguous 140 square degrees of the eROSITA Final Equatorial-Depth survey (eFEDS; Brunner et al. 2022) were observed during the SRG calibration and performance verification phase, between 3 and 7 November 2019. The entire field, centred at RA 136 and Dec +2 (see Fig. 1), was observed to an approximate depth of  $\sim 2.2$  ks ( $\sim 1.2$  ks after correcting for telescope vignetting), corresponding to a limiting flux of  $F_{0.5-2\text{keV}} \sim 6.5 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ . The eFEDS field was chosen from among the extragalactic areas with the richest multi-wavelength coverage visible by eROSITA in November 2019. The observations are just about 50% deeper than anticipated for eRASS:8 at the end of the planned four-year program in the ecliptic equatorial region ( $\sim 1.1 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$ ; Predehl et al. 2021). In other words, the eFEDS exposure corresponds to roughly the 80th percentile of the expected eRASS:8 exposure distribution over the whole sky. eFEDS therefore is a fair representation of what the final eROSITA all-sky survey will be, enabling scientists to face and solve the challenges that will accompany their work for the duration of the survey.

As discussed in detail in Brunner et al. (2022), the X-ray catalogues generated by the analysis of the eFEDS eROSITA data comprise a main catalogue, with 27 910 sources detected above a detection likelihood of 6 in the most sensitive 0.2–2.3 keV band, and a Hard catalogue, containing 246 sources detected above a detection likelihood of 10 in the less sensitive 2.3–5 keV band.

In this paper, we focus on the point-like (i.e. with an extension likelihood  $\text{EXT\_LIKE} = 0^1$ ) X-ray sources in these catalogues (27 369 and 246 for the main and hard sample, respectively)

<sup>1</sup> This parameter is obtained from the task `srctool` of the eSASS software (Brunner et al. 2022).

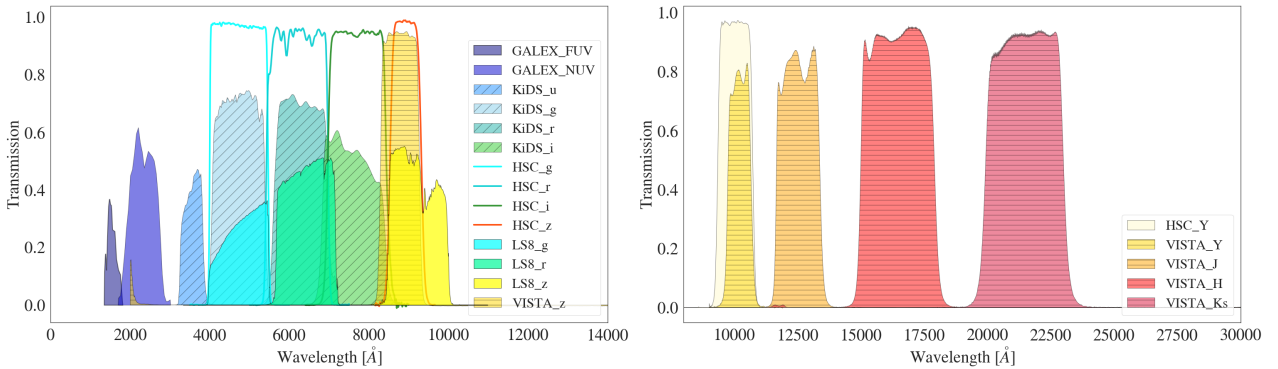
and describe the procedure of (i) reliably identifying multi-wavelength counterparts to the eROSITA sources, (ii) classifying and characterising their properties, and (iii) providing reliable redshift measurements (spectroscopic when available and photometric otherwise). The identification and determination of the reliability of the counterparts, the computation of the photometric redshifts (photo-z), and the characterisation of the sample follow the same procedure for the main and hard samples. For simplicity, we discuss here specifically only the main sample: the two catalogues overlap for 226 of the 246 hard sources, respectively. While we provide the catalogue of counterparts for all the sources in both samples, the properties of the sources in the hard sample are presented and discussed in Nandra et al. (in prep.). The papers about the X-ray spectral analysis (Liu et al. 2022c), variability (Boller et al. 2022; Buchner et al. 2022), X-ray luminosity function (Buchner et al., in prep.; Wolf et al., in prep.), and host properties of AGN in eFEDS (Li et al., in prep.) are all based on the catalogue and/or the method presented in this work. The papers presenting interesting single objects (Brusa et al. 2022; Toba et al. 2022; Wolf et al. 2021), the X-ray properties of WISE sources in eFEDS (Toba et al. 2022), the radio properties of unresolved clusters (Bulbul et al. 2022), and photo-z computed via machine-learning (Nishizawa et al., in prep.) are based on this work as well.

The structure of the paper is as follows: in Sect. 2 we summarise the availability of ancillary data that were used to identify the X-ray counterparts and photo-z estimates. Sections 3 and 4 describe the methods we used to identify the counterparts, while in Sect. 5, the counterparts are finally assigned. In Sect. 6 the counterparts are separated into Galactic and extragalactic sources using morphological, photometric, and proper motion information.

**Table 1.** Photometry available for the counterpart identification and photo-z computation.

| Bands            | Survey      | Depth (AB mag)<br>various indicators | Reference             |
|------------------|-------------|--------------------------------------|-----------------------|
| FUV, NUV         | GALEX       | 19.9, 20.8                           | Bianchi (2014)        |
| $u, g, r, i$     | KiDS        | 24.2, 25.1, 25.0, 23.7               | Kuijken et al. (2019) |
| $g, r, i, z, y$  | HSC         | 26.8, 26.4, 26.4, 25.5, 24.7         | Aihara et al. (2018b) |
| $g, r, i$        | LS8         | 24.0, 23.4, 22.5                     | Dey et al. (2019)     |
| $z, J, H, K$     | KiDS/VIKING | 23.1, 22.3, 22.1, 21.5, 21.2         | Kuijken et al. (2019) |
| $J, K_s$         | VISTA/VHS   | 21.1, 19.8                           | McMahon et al. (2013) |
| $W1, W2, W3, W4$ | LS8/WISE    | 21.0, 20.1, 16.7, 14.5               | Meisner et al. (2019) |

**Notes.** For LS8, the required depth for DESI is listed. For LS8/WISE, the listed depth is taken from Meisner et al. (2019) and is computed using WISE detected sources. Because the photometry we used is forced photometry at the position of optically detected sources, the depth is higher.

**Fig. 2.** Illustration of the bandpass and relative transmission curves of the UV, optical, and near-infrared photometry we used to compute photo-z. For clarity, we do not show the WISE bandpasses.

Because of its size, the field is well populated by stars, AGNs, clusters, and nearby galaxies. Each eROSITA working group has developed independent methods for the identification of sources of interest. In Sects. 5.3 and 7.5, a comparison is made with two main source classes: stellar coronal emitters, and clusters of galaxies. The ultimate goal is to consolidate the counterparts and classify them at the same time. Section 6 presents the multi-wavelength properties of the counterparts and characterises their Galactic or extragalactic nature. Section 7 presents and discusses the photo-z computed with Le PHARE (Ilbert et al. 2006; Arnouts et al. 1999), including a comparison with DNNz (Nishizawa et al., in prep.), an independent method based on machine-learning. Section 8 describes the released data. The basic properties of the point-source eFEDS population based on redshift, photometry, and X-ray flux are presented in Sect. 9. The conclusions in Sect. 10 close the paper, with a forecast of the results and challenges that we will face when working with data from the eROSITA all-sky survey.

The description of the catalogs that we release is provided in the appendix, together with the list of templates we used to compute the photo-z. Throughout the paper, we assume AB magnitudes unless stated otherwise. In order to allow direct comparison with existing works from the literature of X-ray surveys, we adopt a flat  $\Lambda$ CDM cosmology with  $h = H_0/[100 \text{ km s}^{-1} \text{ Mpc}^{-1}] = 0.7$ ,  $\Omega_M = 0.3$ , and  $\Omega_\Lambda = 0.7$ .

## 2. Supporting data

For studies of X-ray sources (taken singularly or as a population), the entire spectral energy distribution (SED) needs to

be constructed and the redshift needs to be determined. Redshift can only rarely be obtained directly from X-ray spectra. It is instead routinely obtained either via optical or near-infrared spectroscopy or via photometric techniques. However, for this to work, the counterparts to the X-ray sources need to be determined first. Deep and homogeneous multi-wavelength data are therefore a prerequisite for any complete population study of an X-ray survey.

The main challenge is that in survey mode, eROSITA has a half-energy width (HEW) of  $26 \text{ arcsec}^2$  (Predehl et al. 2021), which makes the identification of the correct counterparts not at all trivial (e.g. see the review of Naylor et al. 2013; Salvato et al. 2019), especially when we consider in addition that wide-field multi-wavelength homogeneous surveys are very difficult to obtain, with very few exceptions (see the Legacy imaging survey supporting the Dark Energy Spectroscopic Instrument, DESI; Dey et al. 2019). By construction, the eFEDS field is placed in an area that fully encompasses the GAMA09 equatorial field (Driver et al. 2009) and is rich in ancillary photometric and spectroscopic data (Merloni et al., in prep.). We list and describe the surveys we used in more detail below. Table 1 summarises the depth in each filter, and Fig. 2 shows the coverage of the ancillary data in wavelength.

### 2.1. Supporting the associations

The counterparts were identified using the DESI Legacy Imaging Survey DR8 (LS8; Dey et al. 2019) for various reasons.

<sup>2</sup> i.e. comparable to the XMM Slew Survey: <https://www.cosmos.esa.int/web/xmm-newton/xmms12-ug>

First of all, LS8 covers the field homogeneously and has sufficient depth, based on the expected optical properties of the X-ray population (Merloni et al. 2012). In addition, the survey together with *Gaia* also provides the AllWISE tractor (Lang 2014) photometry extracted at the position of the optical sources. Finally, the survey covers 14 000 square degrees of sky, thus providing a sufficient number of sources external to eFEDS that can be used as training and validation samples to test the association (see Sect. 3). The absolute astrometry of the LS8 catalogue is registered to the *Gaia* DR2 astrometric frame, with residuals typically smaller than 30 milliarcseconds<sup>3</sup>. While the catalogued positional uncertainties of individual LS8 sources are often much larger than this systematic limit (especially toward fainter magnitudes), they are still extremely small relative to those of the eFEDS X-ray sources and are set to 0.1'' for the entire survey. Photometry and parallax measures from *Gaia*, which are optimised for point-like sources, are ideal for the identification of the stars in our sample. For this purpose, EDR3<sup>4</sup> has been used (Gaia Collaboration 2020) instead of the *Gaia* DR 2 provided by LS8.

## 2.2. Supporting photometric redshifts

To compute the photo-*z*, we used the following data sets:

**GALEX.** The NASA satellite GALEX has mapped the entire sky in the far- and near-UV (FUV and NUV) between 2003 and 2012, with a typical depth of 19.9 and 20.8 AB magnitude in FUV and NUV, respectively. We used the catalogue catalogue GR6/7 presented in Bianchi (2014) that is available via Vizier.

**Kilo-degree Survey (KiDS).**<sup>5</sup> The survey mapped 1350 deg<sup>2</sup> in *u*, *g*, *r*, *i* bands using VST/OmegaCAM. The same area was also covered by the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING; Edge et al. 2013) in Z, Y, J, H, K. We used the catalogue presented in Kuijken et al. (2019); it has ZYJHK aperture-matched forced photometry to the *ugri* source positions. About 65 deg<sup>2</sup> of sky are shared between KiDS/VIKING and eFEDS.

**HSC S19A.** The Hyper Suprime-Cam (HSC; Miyazaki et al. 2018) Subaru Strategic Program survey (HSC-SSP; Aihara et al. 2018a) is an ongoing optical imaging survey with five broadband filters (*g*-, *r*-, *i*-, *z*-, and *y*-band) and four narrow-band filters (see Aihara et al. 2018b). We used S19A wide data obtained from March 2014 to April 2019, which provide forced photometry for the five bands, with the 5 $\sigma$  limiting magnitudes as listed in Table 1 (see Aihara et al. 2018b, 2019). The astrometric uncertainty is approximately 40 milliarcseconds in rms.

**VISTA/VHS.** The entire southern hemisphere has been observed by VISTA in the near-infrared, and at least for J and Ks, the depth is 30 times the depth of 2MASS (McMahon et al. 2013). We used the DR4 data that are available via Vizier.

**WISE.** The Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) scanned the entire sky in the 3.4, 4.6, 12, and 22  $\mu$ m bands over the course of one year (hereafter W1, W2, W3, and W4). Afterwards, the survey continued with observations in W1 and W2 only. The photometry in W1, W2, W3, and W4 from LS8 includes all five years of publicly available WISE and

NEOWISE reactivation (Meisner et al. 2019). It was measured using the TRACTOR algorithm (Lang 2014) at the position of *grz* detected sources.

## 2.3. Optical spectroscopy

The eFEDS field has previously been observed by several spectroscopic surveys, most notably GAMA, SDSS, WiggleZ, 2SLAQ, and LAMOST. Many of the existing spectra are of high enough quality for us to use them for science applications, in particular, when we just need redshift and basic classification (i.e. deciding whether the source is a star, QSO, or galaxy). However, a careful collation and homogenisation of the existing spectroscopy catalogues was first needed to provide a reliable compendium of these data.

The largest body of spectroscopic redshift information comes from the SDSS survey (York et al. 2000; Gunn et al. 2006; Smee et al. 2013; Abdurro'uf et al. 2022), totalling more than 68k spectra of 61k science targets within the outer bounds of the eFEDS field. We collected archival public data from SDSS phases I-IV (Ahumada et al. 2020), as well as the results of the recent dedicated SPIDERS (Spectroscopic identifications of eROSITA sources) campaign (Comparat et al. 2020, Merloni et al., in prep.), within SDSS-IV (Blanton et al. 2017) following up eFEDS X-ray sources. A small team of the authors visually inspected all of the SDSS 1D spectra that lie in the vicinity of eFEDS X-ray sources, correcting occasional pipeline failures, and grading the spec-*z* onto a common normalised quality (NORMQ) scale between 3 and -1. NORMQ can be interpreted as follows: spec-*z* with NORMQ=3 are those with 'secure' spectroscopic redshifts, those with NORMQ=2 are 'not secure' (although a large fraction are expected to be at the correct redshift), spec-*z* with NORMQ=1 are 'bad' (e.g. low Signal to Noise, S/N, problematic extraction, dropped fibres), and those with NORMQ=-1 are 'blazar candidates'. For completeness, we also retained SDSS-DR16 spectroscopic redshifts in the eFEDS field that do not lie near eFEDS X-ray detections, but only when they satisfied all the following criteria: SN\_MEDIAN\_ALL>2.0, ZWARNING = 0, SPECPRIMARY = 1, and 0 < Z\_ERR < 0.002. An exhaustive description of the SDSS dataset within the eFEDS field will be presented separately by Merloni et al. (in prep.).

We also gathered published spectroscopic redshifts and classifications (hereafter 'spec-*z*') from the literature where they overlap the eFEDS footprint. The detailed breakdown is presented in Table 2. In order to gather spec-*z* from smaller surveys that might only contribute a few redshifts each, we also queried the Simbad database (as of 5 March 2021; Wenger et al. 2000) in the vicinity of the eFEDS X-ray counterpart positions.

For the purposes of this work, we placed greater weight on purity than on completeness. Therefore, where the parent survey catalogues included some metric of quality/reliability, we applied strict criteria to retain only the most secure spec-*z* information. The filtering criteria applied to the original catalogues and the number of spec-*z* considered from each catalogue are listed in Table 2. We assumed that after these quality filtering steps, all the archival spec-*z* are 'secure' (i.e. NORMQ=3), except for Simbad, for which we adopted NORMQ=2. This is meant in this case to be interpreted as 'not yet proven to be secure'.

All these spec-*z* were progressively collated into a single catalogue, with a single redshift and classification per sky position, using a match in coordinates between the coordinates listed for the nine input spectroscopic catalogs. An appropriate search radius (in the range 1–3 arcsec) was chosen according to the

<sup>3</sup> <https://www.legacysurvey.org/dr8/description/#astrometry>

<sup>4</sup> <https://www.cosmos.esa.int/web/gaia/earlydr3>

<sup>5</sup> <http://kids.strw.leidenuniv.nl/>

**Table 2.** Spectroscopic redshift measurements within the eFEDS footprint ( $126 < \text{RA} < 146.2$  deg,  $-3.2 < \text{Dec} < +6.2$  deg).

| Spectroscopic survey | Quality threshold   | $N_{\text{specz}}$ | Data release and reference             |
|----------------------|---|--------------------|--|
| SDSS                 | See text  | 46 837             | Up to DR17; Merloni et al., in prep.   |
| GAMA                 | $\text{NQ} \geq 4$  | 26 318             | DR3; Baldry et al. (2018)              |
| WiggleZ              | $Q \geq 4$  | 13 466             | Final DR; Drinkwater et al. (2018)     |
| 2SLAQ                | $q_{z2S} = 1$   | 953                | v1.2; Croom et al. (2009)              |
| 6dFGS                | $4 \leq q_{cz} \leq 6$                                      | 365                | Final DR; Jones et al. (2009)          |
| 2MRS                 | Non-null velocity   | 152                | v2.4; Huchra et al. (2012)             |
| LAMOST               | $\text{snrr} > 10, z > -1, 0.0 \leq z_{\text{err}} < 0.002$ | 55 866             | DR5v3; Luo et al. (2015)               |
| Gaia RVS             | Non-null velocity   | 15 568             | DR2; Gaia Collaboration (2018)         |
| Simbad               | Non-null redshift   | 3915               | As of 05/03/2021; Wenger et al. (2000) |
| Total unique objects |   | 143 637            |  |

**Notes.**  $N_{\text{specz}}$  is the number of spectroscopic redshift measurements that pass the quality threshold (applied to columns provided in the originating catalogue). For Simbad, the number of entries is limited to objects lying within 3 arcsec of the optical coordinates of counterparts to eFEDS sources. Some astrophysical objects appear in two or more redshift catalogues.

expected positional fidelity and/or fibre sizes associated with each input spectroscopic catalogue. After this de-duplication step, we were left with 143 637 unique entries over the eFEDS field, 108 834 of which are secure (i.e.  $\text{NORMQ} = 3$ ).

### 3. Counterpart identification: Method

Because of the large PSF of the eROSITA telescopes and the small number of photons associated with typical X-ray detections, the  $1\sigma$  rms positional uncertainties of individual X-ray sources can be several arcseconds. Specifically, in eFEDS, the mean positional error is 4.7 and extends above 20 arcsec only for a handful of sources (Brunner et al. 2022). For the expected optical/infrared magnitude distribution of X-ray sources at the depth of the eFEDS (see e.g. Merloni et al. 2012; Menzel et al. 2016), the sky density of the relevant astrophysical source populations is often high. For this reason, the identification of the true associations cannot be determined solely by closest-neighbour searches, as there will be several potential counterparts within the error circle of any given X-ray source. Taking this into account, the identification of the counterparts of eFEDS point-like sources has been performed using two independent methods.

NWAY (Salvato et al. 2019), based on Bayesian statistics, and ASTROMATCH (Ruiz et al. 2018), based on the maximum likelihood ratio (MLR; Sutherland & Saunders 1992), have been specifically developed to identify the correct counterparts to X-ray sources, independently of their Galactic or extragalactic nature. In order to assess the probability (or likelihood) of an object to be the correct counterpart to an eFEDS sources, the two methods first take the separation between the sources, their positional accuracy and the number density of the sources in the ancillary data into account. The difference between the methods resides then in the adoption of specific features that are able to distinguish an X-ray emitter (regardless of its Galactic or extragalactic nature) from a random source in the field. Both methods determined the features (priors) using a representative training sample constructed using secure counterparts to X-ray sources detected in 3XMM. In the case of NWAY, the prior was also determined by comparing the features of the sources in the training sample with the features of the field sources present within 30 arcsec from the 3XMM sources. The disentangling power of the priors was then tested on a blind validation sample of 3415

*Chandra* sources with secure counterparts, where the accuracy of the *Chandra* position was made eFEDS-like (see Sect. 4.1). The detailed description of the construction of training, validation, and respective field sample is presented in Appendix A. Here we provide a short description of NWAY and ASTROMATCH and how their respective priors were determined.

#### 3.1. NWAY enhanced with photometric priors defined via machine-learning

In addition to astrometry, that is, (a) the separation between an X-ray source and a candidate counterpart; (b) the associated positional uncertainties and (c) the number densities of the sources in the two catalogs, the photometry of potential counterparts is valuable information to determine whether they are associated with a given X-ray detection. Traditionally, the likelihood ratio associated with angular distance was multiplied by a factor accounting for the magnitude distributions and the sky density of a population of X-ray sources and background objects (e.g. Brusa et al. 2005, 2007; Luo et al. 2010). In NWAY, this idea was re-formulated in the Bayesian formalism in the following way.

Given some data  $D$ , the posterior association probability  $P(H | D)$  is related to the prior probability  $P(H)$  via the likelihood  $P(D | H)$ ,  $P(H | D) \propto P(H) \times P(D | H)$ .  $P(H)$  is computed from the source densities in each catalogue. If photometric information (or any other feature, in fact) is used, then the likelihood becomes  $P(D | H) = P(D_\phi | H) \times P(D_m | H)$ , where  $D_\phi$  and  $D_m$  refer to the astrometric and photometric information, respectively. For any possible association, the modifying factor  $P(D_m | H)$  is computed from the feature (e.g. magnitude or colour)  $m$  of the counterpart candidate and from the expected distribution of this observable for X-ray sources and field (non-X-ray) sources. We call such factors “priors” to NWAY, as they enter as a priori information in the ultimate matching process. These priors are posteriors previously learned from other data. For further details of the formalism, we refer to Salvato et al. (2019) and the NWAY documentation<sup>6</sup>.

In order to take full advantage of the LS8 ancillary catalogue, we extended this approach for the eFEDS counterpart identification. Instead of using a subset of magnitudes, colours, and their associated distributions, we trained a random forest

<sup>6</sup> <https://github.com/JohannesBuchner/NWAY>

**Table 3.** Final list of LS8 training features used to model the photometric prior in NWAY.

| Feature                   | Description                                       |
|---------------------------|---|
| flux_*/mw_transmission_*  | Dereddened flux in $g, r, z, W1, W2$              |
| gaia_phot_*_mean_mag      | Original <i>Gaia</i> phot. in $G, G_{bp}, G_{rp}$ |
| snr_*                     | S/N for $g, r, z, W1, W2, G, G_{bp}, G_{rp}$      |
| $\sqrt{pmra^2 + pmdec^2}$ | <i>Gaia</i> proper motion                         |
| parallax                  | <i>Gaia</i> parallax                              |
| $g-r, r-z, z-W1, r-W2$    | Dereddened colours                                |

classifier (sklearn implementation; Pedregosa et al. 2011) on a large number of features to reliably map the available Legacy DR8 information to real X-ray sources and real field objects, the details of which are described in the next subsection. The trained classifier is then used to predict the probability of all counterpart candidates to be X-ray emitting, also taking the spatial information into account, as described in Sect. 3.1.2. This probability is directly used to compute  $P(D_m | H)$ . In the following section, we describe the definition of the features in the training sample.

### 3.1.1. Random forest prior: Training and performance

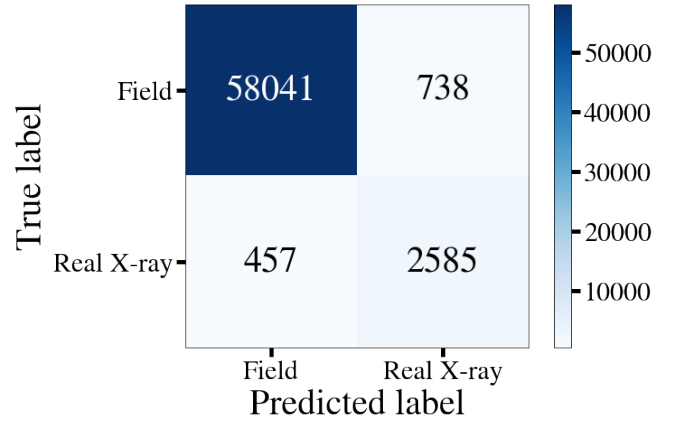
From the 3XMM training sample described in Appendix A, we extracted a set of photometric and astrometric features. The training features are listed and described in Table 3. X-ray sources are flagged as target class 1 and field objects as target class 0. About 15% of the 3XMM training samples (61821 sources) were extracted randomly for testing purposes and were not further considered in the training procedure. The baseline model is composed of 200 trees, allowing decision split points if at least eight samples are left in each branch. All of the 22 features can be used to build the decision tree, which makes use of bootstrap samples of the training set.

By construction, the training sample is highly imbalanced, since the field objects strongly outnumber the X-ray sources. We therefore opted for a weighting scheme that automatically adjusted weights of training examples for the class imbalance.

The trained model was evaluated on the test set, resulting in the confusion matrix presented in Fig. 3. We note that the cut in the class prediction for the presented confusion matrix is made at  $p_{X\text{-ray}} = 0.50$ , where  $p_{X\text{-ray}}$  is the predicted probability that a counterpart candidate is X-ray emitting. Since NWAY uses the continuous predicted probability as modifying factor for the likelihood  $P(D | H)$ , real counterparts with rare or untypical photometric features, that is, with  $p_{X\text{-ray}} \lesssim 0.50$ , may still be selected by the algorithm if the astrometric configuration favours them. We obtained a high recall fraction of  $2585/(2585 + 457) = 85\%$ , while the fractional leakage of contaminating field objects remained low:  $738/(738 + 58041) = 1\%$ . We note that  $p_{X\text{-ray}}$  was only computed from the photometric and proper motion properties of the LS8 sources. In particular, it does not depend on coordinates and positional uncertainties. As discussed in the previous section, this allows us to split the likelihood of a match into independent astrometric and photometric terms:  $P(D | H) = P(D_\phi | H) \times P(D_m | H)$ .  $P(D_m | H)$ , the photometric term, is directly related to  $p_{X\text{-ray}}$ .

### 3.1.2. NWAY association run

Using the trained model, we predicted  $p_{X\text{-ray}}$  for all LS8 sources in the eFEDS field. We then ran the NWAY matching procedure

**Fig. 3.** Confusion matrix resulting from the random forest prediction on an independent test set. X-ray sources are labelled “real X-ray”, and field objects are labelled “field”. Numbers on the right downward diagonal correspond to correctly predicted classes. At this step, the separation between the sources and the X-ray position is not considered.

using the ratio  $p_{X\text{-ray}}/(1 - p_{X\text{-ray}})$  for  $P(D_m | H)$ . This was done by adding  $p_{X\text{-ray}}$  as a column to the LS8 catalogue and activating it as a prior column in NWAY with the `-mag` option. We set a radius of 30 arcsec from each eFEDS X-ray source, considering all LS8 sources within this radius. This may appear to be a relatively large maximal separation, given the mean eFEDS positional error of 4.5 arcsec; however, we wish to account also for the largest positional uncertainties of a few objects in the eFEDS source catalogue (see Brunner et al. 2022) and the use of a large search radius minimises the probability of missing counterparts that are widely separated from the X-ray centroid position. The sky coverage of eFEDS and LS8 is  $140 \text{ deg}^2$  and  $N_{\text{eFEDS}} \times \pi \times (30'')^2 - A_{\text{overlap}}$ , respectively, where  $N_{\text{eFEDS}}$  is the number of eFEDS sources (point-like or extended) and  $A_{\text{overlap}}$  is the overlap area of neighbouring search windows around the X-ray sources. As described in the appendix of Salvato et al. (2019), the area coverages were used to compute the number densities, which in turn were used to compute the probability for an eFEDS source to have a counterpart ( $p_{\text{any}}$ ) and the probability for each source in LS8 to be the correct counterpart ( $p_{\text{i}}$ ). These two quantities were then used to assign a counterpart to an eFEDS source: while the LS8 source with the highest  $p_{\text{i}}$  was considered to be the best available counterpart, we used the  $p_{\text{any}}$  value to decide whether the identification of the counterpart was reliable (see Sect. 4.2 for details).

### 3.2. MLR approach

The maximum likelihood ratio (MLR) statistic for the correct pairing of sources from multiple catalogues was introduced in the seminal work of Sutherland & Saunders (1992) and is widely used, although mostly to pair only two catalogs. For sources detected at two different wavebands and separated by angular distance  $r$  on the plane of the sky, the likelihood ratio provides a measure of the probability that the two sources are true counterparts normalised by the probability that they are random alignments. Quantitatively, this is estimated as

$$LR = \frac{q(\vec{m}) \cdot f(r)}{n(\vec{m})}, \quad (1)$$

where  $q(\vec{m})$  is the prior knowledge about the properties of the true associations, such as the distribution of their apparent

magnitudes at given spectral window, their colours, and/or the spatial extent of the observed light in a given waveband. The collection of all possible source properties for which a prior probability can be estimated is represented by the vector  $\vec{m}$ . The quantity  $n(\vec{m})$  is the sky density of all known source populations in the parameter space of  $\vec{m}$ . It measures the expected contamination rate from background/foreground sources that are randomly projected on the sky within a distance  $r$  off a given position. The probability that the true associations are separated by the distance  $r$  is measured by the quantity  $f(r)$ . This depends on the positional uncertainties of the matched catalogues.

For the MLR applied to the eFEDS X-ray sources, a multi-dimensional prior was used that combines knowledge of the optical and mid-infrared colours/magnitudes of X-ray sources as well as their optical extent, that is, point-like versus extended.

The version of MLR we applied to the eFEDS work is based on the ASTROMATCH<sup>7</sup> implementation. This tool has been specifically designed to deal with the complexity of wide-area surveys that contain a very large number of sources. The HEALPix multi-order coverage map (MOC<sup>8</sup>) technology was used to describe the footprint of a catalogue of astrophysical sources. The KD-tree library as implemented in the ASTROPY package (Astropy Collaboration 2013, 2018) was used to accelerate spatial searches of potential counterparts within a radius  $r$  of a given sky position. The core ASTROMATCH functionality was expanded to enable the use of multi-dimensional priors. The version of ASTROMATCH adopted in this work is therefore a fork (github.com/ageorgakakis/astromatch) of the main development branch.

Like for NWAY, the optical counterparts were investigated out to a maximum radius of 30 arcsec. We assumed that the positional uncertainties of the X-ray and optical catalogues follow a normal distribution. The quantity  $f(r)$  is therefore represented by a Gaussian with  $\sigma$  parameter estimated as the sum in quadrature of the X-ray and optical positional uncertainties.

The priors were generated using the training sample defined in Appendix A.1. The LS8 photometric properties of the sources in that sample were explored to identify parameter spaces in which they are separate from the general LS8 field population. After some experimentation, we opted for the following three independent priors:

(1) A space that includes the WISE colour  $W1 - W2$ , the WISE magnitude  $W2$ , and the optical extent of a source. For the latter, we used the LS8 parameter TYPE, which provides information about the optical morphology of sources. In our application we only differentiated between optically unresolved (TYPE = "PSF") and optically extended (TYPE  $\neq$  "PSF") populations;

(2) A space that includes the optical/WISE colour  $r - W2$ , the optical magnitude  $g$ , and the optical extent of a source. For the latter, we used the Legacy-DR8 parameter TYPE, as explained above;

(3) The distribution of the *Gaia*  $G$  magnitudes listed in the LS8 catalogues. This is to identify X-ray sources associated with very bright counterparts;

The distribution of the training sample sources in the parameter spaces above was used to define two three-dimensional priors and one one-dimensional independent prior. These were provided as input to the ASTROMATCH code, together with the distribution of the sources in the field population when the association for eFEDS was computed. For a given eFEDS source, all

the potential associations within the search radius of 30 arcsec were identified. Each of them was assigned one LR value for each of the three priors using Eq. (1). The LS8 source with the highest value of LR from one of the three priors was considered to be the counterpart.

## 4. Comparing NWAY and ASTROMATCH on a validation sample

In order to compare completeness and purity of NWAY and ASTROMATCH, the same setting as we adopted to identify the counterparts to eFEDS was used to determine the best counterparts to a blind validation sample of 3415 counterparts to *Chandra* sources (see Appendix A). This validation sample was used as a truth table to test the performance of NWAY and ASTROMATCH in finding counterparts and to define the `p_any` and `LR_BEST` thresholds above which a counterpart is considered secure.

### 4.1. eROSITA-like validation sample

The *Chandra* sources were assigned eROSITA positional errors by randomly sampling from the astrometric uncertainties listed in the core eFEDS source catalogue. We accounted for the flux dependence of these uncertainties by matching any given *Chandra* source with a certain flux from 0.5–2 keV to only those eFEDS sources with a similar 0.6–2.3 keV flux within a margin of 0.5 dex. The flux transformation between the *Chandra* and eFEDS spectral bands is small, about 2% for a power-law spectral energy distribution with  $\Gamma = 1.9$ , for instance, and was ignored. The positional uncertainty,  $\sigma$ , assigned to each of the *Chandra* sources, can be split into a right-ascension and a declination component. It was assumed that these two uncertainties are equal, and therefore,  $\delta\text{RA} = \delta\text{Dec} = \sigma/\sqrt{2}$ . Under the assumption that both the  $\delta\text{RA}$  and  $\delta\text{Dec}$  are normally distributed, the total radial positional uncertainty follows the Rayleigh distribution with a scale parameter  $\sigma$ .

Instead of directly using the assigned  $\sigma$  as the astrometric error to be applied to the *Chandra* positions to make them resemble the eFEDS astrometric accuracy, we preferred to add further randomness to the experiment. For each *Chandra* source, the assigned  $\sigma$  was treated as the scale factor of the Rayleigh distribution and a deviate was drawn that represented the positional error. This was applied to the sky coordinates of the optical counterpart of the *Chandra* source, and the new offset position was taken as the centroid of the X-ray source in the case of an eFEDS-like observation.

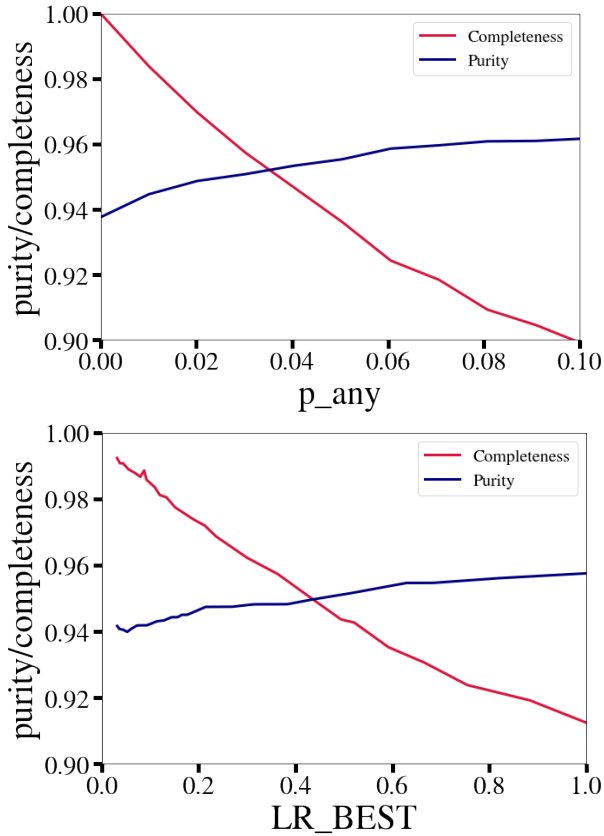
### 4.2. Probability thresholds definition

The LS8 counterparts to the *Chandra* eFEDS-like sources were identified by NWAY and ASTROMATCH using the same setup as we adopted for the real eFEDS observation. The resulting catalogue of best counterparts was matched with the true associations, providing a direct comparison between the methods, and, at the same time, providing a measure of the false-positive identification rate of the eFEDS counterpart catalogue.

First, we compared the primary identifications returned by NWAY and ASTROMATCH to true identifications stored in the validation sample. NWAY and ASTROMATCH correctly identified 3216 out of 3394 (95%) and 3024 out of 3394 (89%) sources, respectively. NWAY has a higher success rate. Additionally, NWAY has a smaller fraction of sources with a second

<sup>7</sup> <https://github.com/ruizca/astromatch>

<sup>8</sup> <https://www.ivoa.net/documents/MOC/>



**Fig. 4.** Purity (solid blue line) and completeness (solid red line) as a function of  $p\_any$  for the association of the CSC2 eROSITA-like validation sample made with NWAY (*top panel*) and as a function of LR\_BEST for the ASTROMATCH (*bottom panel*).

possible counterpart (115 sources against 367). Another way to look at the results is to compare purity and completeness for the two methods. At any given value of  $p\_any$ /LR\_BEST, we defined as purity the fraction of sources with a correct identification. In addition, we defined as completeness the fraction of sources for which we were able to assign a counterpart (see Fig. 4). Both methods have very high purity and completeness. NWAY provides a sample that is purer, consistent with the fact that very few sources have a second possible counterpart, in addition to the correct one. Combined with the success rate, this makes NWAY the more robust method for determining the counterparts. Its strength comes first of all from the capability to account for complicated priors involving multiple features (essentially resembling an entire SED, together with other physical properties), from different catalogs at the same time. Furthermore, the Bayesian statistics upon which NWAY is based also allows accounting for sources that are lacking one or more of the features.

Similarly to what is traditionally done in maximum likelihood (see e.g. Brusa et al. 2007), the intersection between the completeness and purity can be used to define a threshold above/below which the counterparts is considered reliable. This corresponds to 0.035 for  $p\_any$  and 0.45 for LR\_BEST.

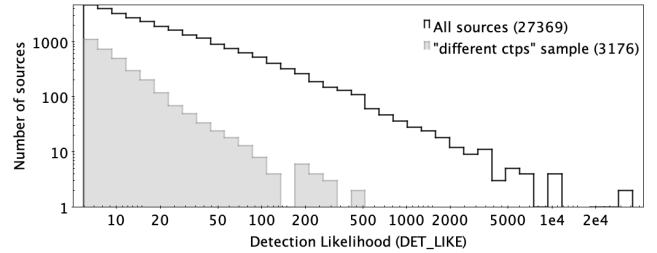
## 5. Determination of the counterparts to eFEDS sources

While for the large majority of the cases the two methods select the same counterparts, there are cases where they do not agree

**Table 4.** Comparison of matches between NWAY and ASTROMATCH.

| Sample          | Number | Counterparts |              |
|-----------------|--------|--------------|--------------|
|                 |        | “Same”       | “Different”  |
| DET_LIKE_0 > 6  | 27 369 | 24 193       | 3176 (11.6%) |
| DET_LIKE_0 > 8  | 21 410 | 19 162       | 1795 (8.4%)  |
| DET_LIKE_0 > 10 | 17 574 | 16 435       | 1136 (6.5%)  |

**Notes.** In the last column, the fraction of the “different ctps” with respect to the whole sample is also reported.



**Fig. 5.** Number of sources as a function of detection likelihood for the entire sample of eFEDS main catalogue sources (black histogram) and for the sources for which NWAY and ASTROMATCH indicate different counterparts (grey shaded histogram).

or where they identify multiple likely associations. In the following we describe the procedure that we adopted for the final assignment of the counterparts.

Then, after the consolidation of the counterpart, we describe a further test for consistency that was done by comparing the results of the association with an independent method, Ham-Star (Schneider et al. 2022), which is tuned to identify Galactic coronal X-ray emitters (Sect. 5.3). The same process was then repeated for the 246 sources in the eFEDS hard point-source catalog. From now on, all numbers and descriptions are given for the main sample unless specified otherwise.

### 5.1. Comparison of counterparts from NWAY and ASTROMATCH

For 24 193 out of 27 369 (88.4%) eFEDS point-like sources in the main sample, NWAY and ASTROMATCH point at the same counterpart. They disagree for 3176 (11.6%) of the cases. The numbers are quoted at this stage regardless of the  $p\_any$  or LR\_BEST thresholds, which are used instead below in order to assign a flag for the quality of the proposed counterpart.

Table 4 summarises the number of eFEDS sources with the agreement/disagreement between the two methods as a function of detection likelihood of the X-ray source. Sources with low detection likelihood values have larger X-ray positional errors on average, and a larger number of spurious sources is expected from simulations (Brunner et al. 2022; Liu et al. 2022c). It is therefore not surprising that the largest discrepancies are observed at the lowest detection likelihoods (Fig. 5). The disagreement drops from 11.6 to 6.5% when we consider only eFEDS sources with DET\_LIKE greater than 10, suggesting that at low detection likelihood, a fraction of eFEDS sources might be spurious detections where NWAY and ASTROMATCH assign a different field source. The notion that these are field sources is also supported by the fact that for about 50% of eFEDS sources with DET\_LIKE below 10 and with different counterparts, both  $p\_any$  and LR\_BEST are below the threshold.



**Table 5.** Counterpart quality summary.

| “Same ctps”, single solutions – 23 622 (86.3%) |                                |                                |
|--|--------------------------------|--------------------------------|
|  | p_any<br>>0.035                | p_any<br><0.035                |
| LR_best > 0.45                                 | <b>20 873 (0.763)</b>          | <b>561 (0.020)</b>             |
| LR_best < 0.45                                 | <b>818 (0.030)</b>             | <b>1370 (0.050)</b>            |
| “Same ctps”, multiple solutions – 571 (2.1%)   |                                |                                |
|  | p_any<br>>0.035                | p_any<br><0.035                |
| LR_best > 0.45                                 | <b>505 (0.018)</b>             | <b>7 (3 × 10<sup>-4</sup>)</b> |
| LR_best < 0.45                                 | <b>59 (0.002)</b>              | –                              |
| “Different ctps” – 3176 (11.6%)                |                                |                                |
|  | p_any<br>>0.035                | p_any<br><0.035                |
| LR_best > 0.45                                 | <b>1243 (0.045)</b>            | <b>226 (0.008)</b>             |
| LR_best < 0.45                                 | <b>478 (0.017)</b>             | <b>1225 (0.045)</b>            |
| HamStar <sup>(1)</sup>                         | <b>4 (2 × 10<sup>-4</sup>)</b> | –                              |

**Notes.** Comparison of matches with NWAY and ASTROMATCH as a function of their respective thresholds, split between same counterparts (for cases of single and multiple counterparts) and different counterpart classes; in parentheses, we list the fractions of the total sample. The numbers in each box are colour-coded by their CTP\_quality value (see text for more details): thick black = 4; thick blue = 3; thick cyan = 2; thick purple = 1, and thick red = 0. <sup>(1)</sup>Objects for which HamStar would indicate a different counterpart with  $p_{\text{stellar}} > 0.95$  are by definition sources with CTP\_quality = 2 (see text for details).

Table 5 summarises the comparison between the two methods and also takes the reliability of the associations into account. In this table we further split the sample with the same counterparts (“same ctps” for brevity) into two subsamples: one for which the proposed counterparts are the only associations suggested by both methods (“single solutions”; 86.3% of the entire sample), and one for which, although both methods point to the same associations, at least an additional counterpart at lower significance exists from at least one method (“multiple solutions”; 2.1% of the entire sample).

The different priors and the different methods we used to assign the counterparts explain the selection of different counterparts in the different ctps sample. ASTROMATCH uses three priors, but they are each used independently, and for any given eFEDS source, the counterpart is assigned by the prior with the higher probability. Instead, NWAY uses all the features at the same time, and the best counterpart is the one that mimics best the training sample in a multidimensional space. We consider this second method more reliable, and for this reason, we decided to always list as primary the counterpart suggested by NWAY, unless LR\_BEST is above the threshold and p\_any is not.

Interestingly, we note that in the different ctps sample, the primary counterpart assigned by one method is the secondary counterpart assigned by the other for about 25% of the cases.

### 5.2. Assigning a quality to the proposed counterparts

As a consequence of the discussion above, each counterpart in the catalogue was flagged as follows ([number] refers to the number of sources in the category):

- CTP\_quality = 4: when NWAY and ASTROMATCH agree on the counterpart, and both p\_any and LR\_BEST are above threshold [20 873 sources; black in Table 5];

- CTP\_quality = 3: when NWAY and ASTROMATCH agree on the counterpart, but only one of the methods assigns the counterpart with a probability above the threshold [1379 sources; blue in Table 5];

- CTP\_quality = 2: when there is more than one possible reliable counterpart. This includes a) all the sources in the different ctps sample with at least one probability above the threshold, and b) the sources in the same ctps sample with possible secondary solutions [2522 sources in total; cyan in Table 5]. Because of the low spatial resolution of eROSITA, this last case implies that both sources contribute to the X-ray flux. A supplementary catalogue with the properties of the secondary counterparts for these 2522 sources is also released (Sect. 8).

- CTP\_quality = 1: when NWAY and ASTROMATCH agree on the counterpart, but both p\_any and LR\_BEST are below the threshold [1370 sources; purple in Table 5]; a probability below the threshold does not necessarily imply an incorrect counterpart. It might also indicate that the counterpart is correct, but its features do not sufficiently mimic those in the training sample.

- CTP\_quality = 0: when NWAY and ASTROMATCH indicate different counterparts and both p\_any and LR\_BEST are below the threshold [1225 sources; red in Table 5].

Counterparts with quality 4, 3, and 2 are considered reliable (90.5% of the main sample and 93.9% of the hard sample), while sources with quality 1 or 0 are considered unreliable (9.5% of the main sample and 6.1% of the hard sample).

### 5.3. Comparison with an independent association method tuned to stars: HamStar

The content of the eFEDS point-source catalogue was also analysed in order to specifically identify stellar coronal X-ray emitters with sufficiently well-defined properties. This method, called HamStar in the following, is based on the properties expected for this type of star; the details are presented in Schneider et al. (2022). In short, HamStar performs a binary classification between stellar coronal emitters and other objects. This classification is based on the concept of eligible stellar counterparts, that is, the match catalogue contained only stellar objects that may reasonably be responsible for the X-ray sources. Specifically, the parent sample that HamStar used included only sources from *Gaia* EDR3 that (a) are brighter than 19th magnitude in *G* band (implied by the stellar saturation limit of  $L_X/L_{\text{bol}} \lesssim 10^{-3}$  and the depth of eFEDS); (b) have accurate magnitudes in all three *Gaia* photometric bands (to apply colour-dependent corrections); (c) have a parallax value at least three times higher than the parallax error (to select only genuine stars).

Then, a positional match between sources in eFEDS and the eligible stellar candidates was made, considering all sources within  $5\sigma$  of the positional uncertainty of the eFEDS source as possible stellar counterparts. Finally, the matching probabilities of all possible counterparts were adjusted based on the value of the two-dimensional Bayes map at the counterpart Bp-Rp colour and ratio of X-ray to *G*-band flux. Based on the HamStar algorithm, 2060 eFEDS sources are expected to be stellar (Schneider et al. 2022). The vast majority of them have a unique *Gaia* counterpart, and only 83 eFEDS sources have two possible counterparts.

Of the 2060 eFEDS sources with a counterpart from HamStar, a counterpart for 1883 is identified here that is less than 2 arcsec from the counterpart proposed by Hamstar. We assume

that this is the same source. We visually inspected the cutouts of the 29 sources for which the separation between the counterpart proposed by HamStar and this work is between 2 and 3 arcsec, and concluded that the counterparts are the same for 9 sources, but the sources are heavily saturated in LS8 so that the coordinates are not sufficiently precise. This corresponds to an 92% agreement; incidentally, this value corresponds almost exactly to the expected reliability and completeness of HamStar (Schneider et al. 2022). All these sources are then classified as “secure Galactic” in Sect. 6.

HamStar applies well-understood X-ray-to-optical properties of stars to a well-defined subsample of *Gaia* sources. On the other hand, the training samples used by NWAY and ASTROMATCH include various classes of X-ray emitters: stars and compact objects, AGN, and galaxies, including the bright ones at the centre of clusters (BCG). We considered the prior defined by NWAY and ASTROMATCH to be more representative of the population of X-ray emitters in general and decided to keep the counterpart assigned in the previous section rather than changing counterparts for the 177 sources for which the methods indicate different counterparts. However, we degraded the CTP\_quality because an alternative solution might apply. Interestingly, only 4 out of 211 sources were considered secure, with CTP\_quality = 3. All other counterparts had an already low CTP\_quality.

#### 5.4. Separation and magnitude distribution of the counterparts

For 24 427 out of 24 774 (98.5%) of the sources with CTP\_quality  $\geq 2$ , the separation between the X-ray position and the assigned LS8 counterpart is smaller than 15 arcsec, with a mean of 4.3 arcsec. As might be expected, there is a trend for larger average X-ray-optical separations at lower values of DET\_LIKE; lower detection likelihood sources typically have larger X-ray positional uncertainty (see Brunner et al. 2022). The distribution of the observed X-OIR separations normalised by the X-ray positional uncertainty is shown in Fig. 6 as a function of the  $r$  magnitude of the counterpart. The distribution is broadly comparable to the expectation of a Rayleigh distribution with a scale factor = 1.

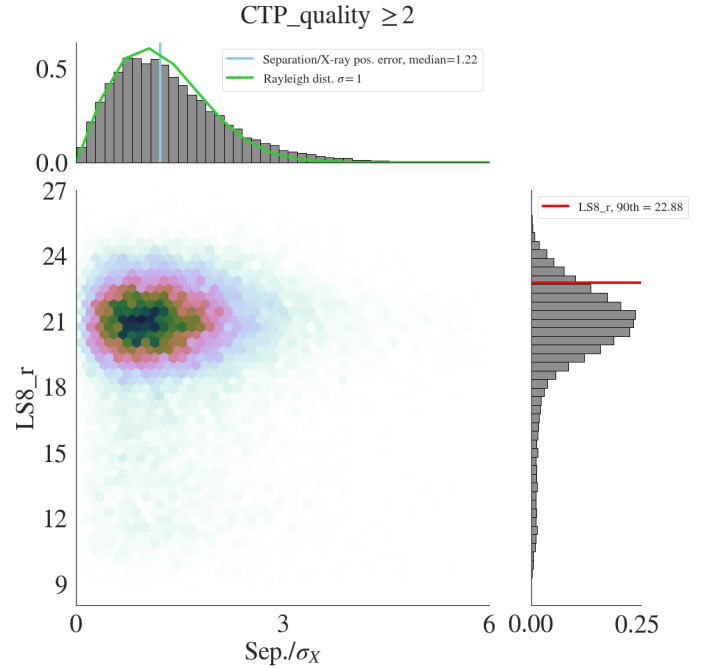
In Fig. 7 we show the distribution of the sample in X-ray flux versus optical magnitude space. The sample is subdivided into objects with more secure counterparts (CTP\_quality  $\geq 2$ ) and objects with less reliable counterparts (CTP\_quality  $\leq 1$ ). The less reliable counterparts tend to have fainter optical magnitudes for a given X-ray flux than the more secure counterparts.

## 6. Source characterisation and classification

After the identification of the counterparts, the different classes of objects need to be separated to understand physical processes and populations. The most important separation is between extragalactic sources (galaxies, AGN, and QSOs) and galactic sources (stars, compact objects, etc.). In the following, we describe how we classified the sources and how the validation tests were performed.

### 6.1. Galactic and extragalactic sources

In order to classify sources in the most reliable way, we used a combination of methods and various information: spectroscopy, parallax measurements from *Gaia*, colours, and morphology from imaging surveys. None of the methods is infallible because



**Fig. 6.** Separation between X-ray position and the selected counterpart normalised by the one-dimensional positional error of the X-ray source as a function of its  $r$ -band magnitude for the sources with secure counterparts (CTP\_quality  $\geq 2$ ). The hexagons are colour-coded linearly according to the counts in the specific bins. The marginal histograms have a linear y-axis. The 90th percentile of the  $r$ -band magnitude distribution (22.88) and the median of the normalised X-OIR separation (1.22) are also reported in the marginal histograms. The  $1\sigma$  Rayleigh distribution expected for the normalised separations is overplotted in green.

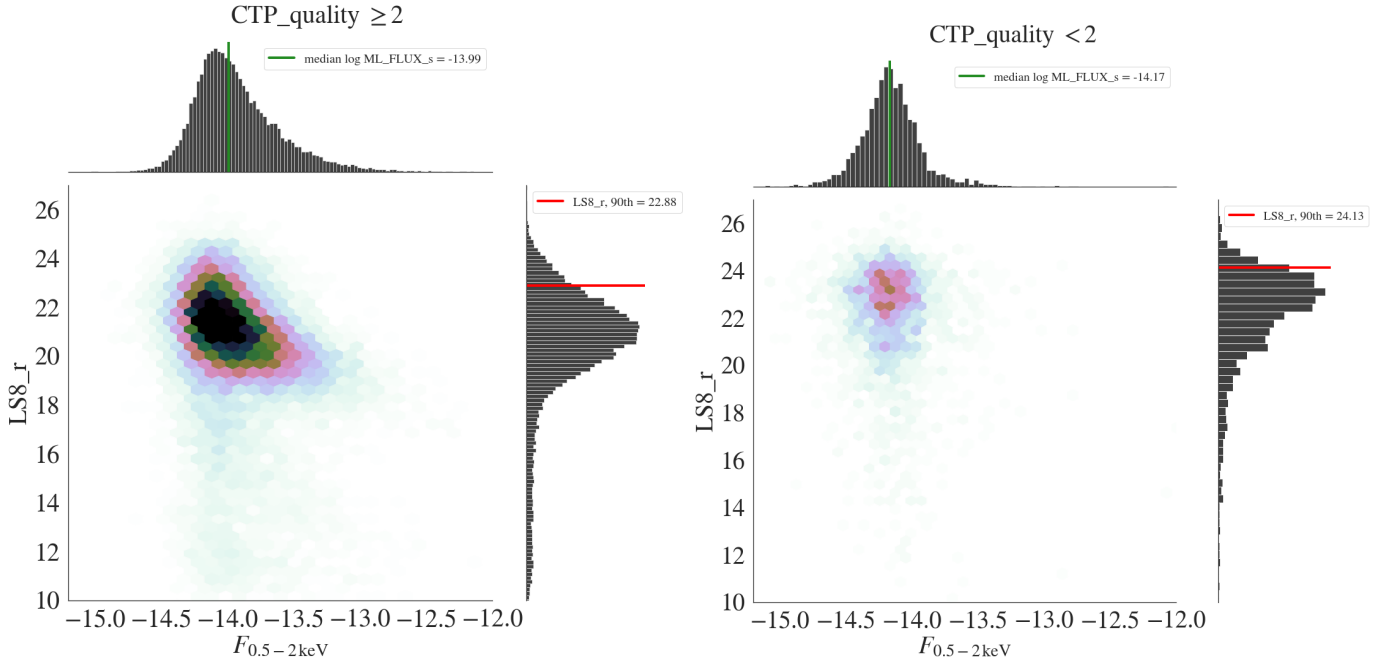
they all depend on the quality of the data (e.g. S/N for spectra, depth, and image resolution) and because of the degeneracy in colour-redshift space for many of the sources. We therefore adopted a multi-step approach: at each step, we extracted from the pool of sources those that can be classified with high reliability as either extragalactic or Galactic. Figure 8 shows an illustration of the decision tree we adopted for the classification, together with the number of sources in each the classes. The procedure is described below in detail.

We first applied the classification based on spectroscopy or high parallax. These can be considered primary methods as they are highly pure, but certainly not complete. The sources thus classified were defined as “secure Galactic” or “secure extragalactic”. Briefly, we defined as secure extragalactic all sources with spectroscopic redshift  $>0.002$  and NORMQ = 3 (step 1 in Fig. 8) and as secure Galactic all sources satisfying at least one of the criteria spectroscopic redshift  $<0.002$  and NORMQ = 3, significant parallax from *Gaia* EDR3 (above  $3\sigma$ ), or agreement with HamStar counterparts (step 2).

Next, we extracted those of the sources that were still in the pool that appeared extended in the optical images. Depending on whether photometry from HSC was available, a source was defined as extended (EXT) if it satisfied

$$\Delta\text{mag} = \text{mag}_{\text{Kron}} - \text{mag}_{\text{psf}} > 0.1 \quad (2)$$

simultaneously in  $g, r, i, z$  from HSC imaging data (e.g. Palanque-Delabrouille et al. 2011), or, when no photometry from HSC was available (either because the source is outside the field or



**Fig. 7.** Magnitude distribution of the counterpart vs. the X-ray flux for sources with  $CTP\_quality \geq 2$  (left) and  $CTP\_quality < 2$  (right). A cut at  $\log ML\_FLUX\_s > -20$  has been applied. The magnitude distribution is clearly different. The green lines represent the median 0.5–2 keV flux (–13.99 and –14.17), while the red lines mark the 90th percentile of the  $r$ -band magnitude distribution (22.88 and 24.13).

because of saturated photometry),

$$LS8\_TYPE \neq PSF. \quad (3)$$

The EXT sources were then flagged as “likely extragalactic” (step 3). This is considered a secondary classifier because in poor seeing conditions, for example, point-like sources (or stellar binary systems) would also be misclassified as extended (see the discussion presented in Hsu et al. 2014).

The sources classified as “secure” were then projected in the LS8  $z$ - $W1$  versus  $g$ - $r$  plane (see inset in the left panel of Fig. 9), following Ruiz et al. (2018). There, we empirically defined a line separator, described as

$$z - W1 - 0.8 * (g - r) + 1.2 = 0, \quad (4)$$

which provides a sharp separation between secure Galactic and extragalactic sources; a negligible fraction of secure extragalactic sources lies below the separator (left panel of Fig. 9). Then, for all the sources still in the pool and with available photometry from LS8 (step 4), we classified the sources below the line as “likely Galactic” (step 5). The remaining sources in the pool with available LS8 photometry were classified as “likely Galactic/extragalactic” (step 6) depending on whether they fell below or above the line in the  $W1$  versus X-ray flux plane (see inset in the right panel of Fig. 9), as defined in Salvato et al. (2019),

$$W1 + 1.625 * \log(F_{0.5-2keV}) + 6.101 = 0, \quad (5)$$

with  $W1$  in the Vega system and X-ray flux in cgs. Originally, a similar line separator was introduced by Maccacaro et al. (1988) using X-ray and optical bands. This was tested over time at different X-ray flux depths or at different wavelengths (e.g. near-infrared; see Civano et al. 2012). This new line separates X-ray bright AGN from X-ray faint stars, and was constructed by combining data from the deep COSMOS *Chandra* Legacy survey (Marchesi et al. 2016) and ROSAT/2RXS (Boller et al.

2016; Salvato et al. 2019). It can be considered a good separator only after the extended nearby extragalactic sources are taken into account (see right panel of Fig. 9). It has the advantage of generality, as the  $W1$  photometry and the X-ray fluxes are available virtually for all the eFEDS sources. Finally, we assumed for the sources without complete information from LS8 that they are extragalactic (step 7), unless they are below the  $W1$ -X line defined in Eq. (5) (step 8).

In this manner, a simple but reliable four-way classification scheme (secure/likely Galactic/extragalactic) was achieved. The final distribution of the four classes of sources in the  $g$ - $r$ - $z$ - $W1$  vs.  $W1$ -X planes is shown in Fig. 10. The two line separators identify four wedges, two of which can be used to define almost 100% pure subsamples of Galactic/extragalactic X-ray selected sources. The four wedges are described below.

- Top left: 724 sources, out of which 637 (87.9%) are Galactic (463 and 428, respectively, only considering sources with reliable counterparts,  $CTP\_quality \geq 2$ ).

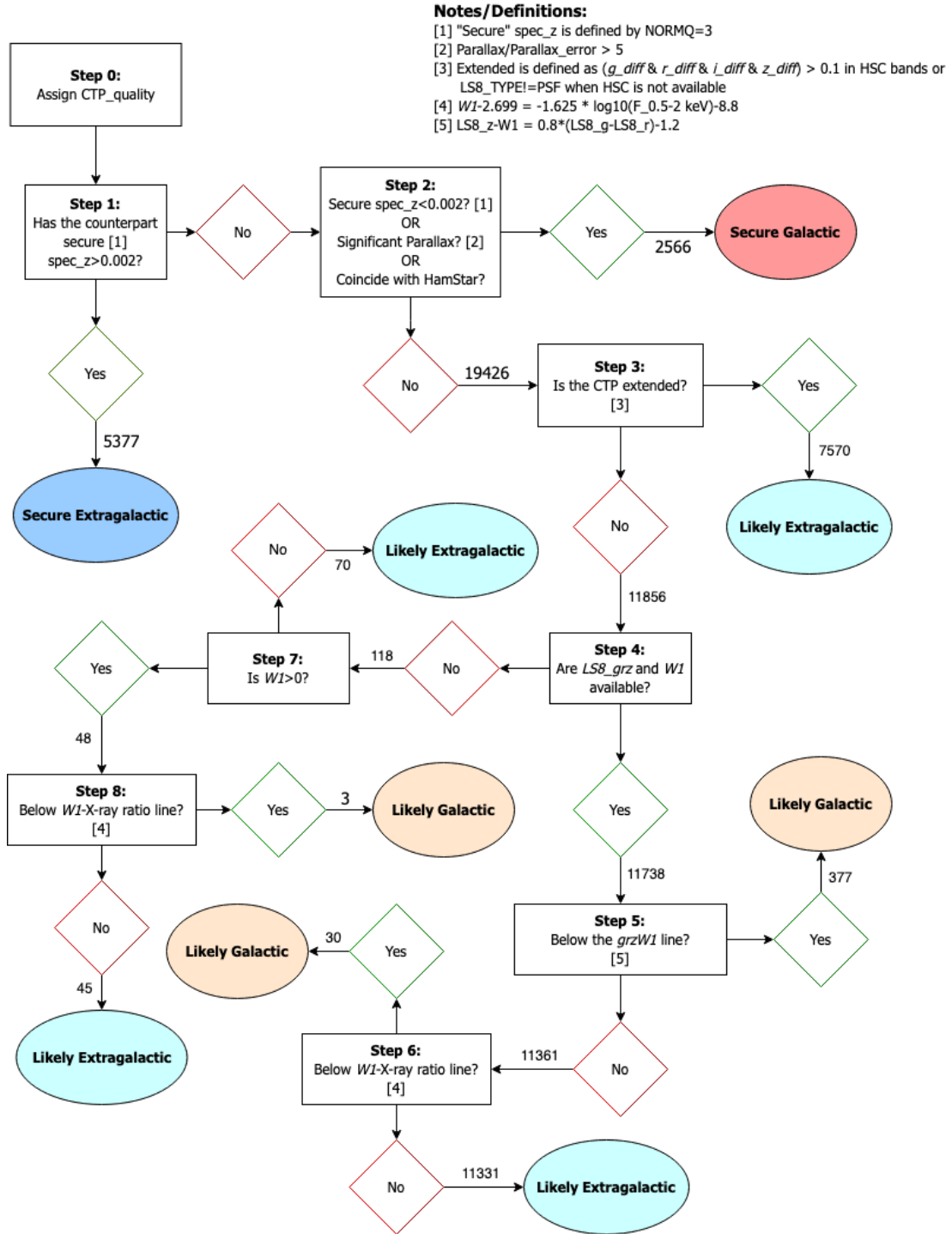
- Top right: 23 874 sources, out of which 23809 (99.7%) are extragalactic (21 711 and 21647 for  $CTP\_quality \geq 2$ ).

- Bottom left: 1391 sources, out of which 1373 (98.7%) are Galactic (1337 and 1319 for  $CTP\_quality \geq 2$ ).

- Bottom right: 1380 sources, out of which 479 (34.7%) are extragalactic (1263 and 379 for  $CTP\_quality \geq 2$ ).

It is important to recall that the order of the steps taken in the decision tree is crucial to limit the misclassification of the sources as much as possible. For example, the use of spectroscopic redshift in the first step allowed us to identify the bright and nearby extragalactic sources that would have been misclassified as Galactic in the  $W1$ -X plane. Similarly, the adoption of the high parallax from *Gaia* allowed us to identify secure galactic sources that would have been misclassified as extragalactic in the  $z$ - $W1$  versus  $g$ - $r$  plane.

In summary, the eFEDS main sample comprises 24 393 sources classified as extragalactic (5377 secure and 19 016 likely)



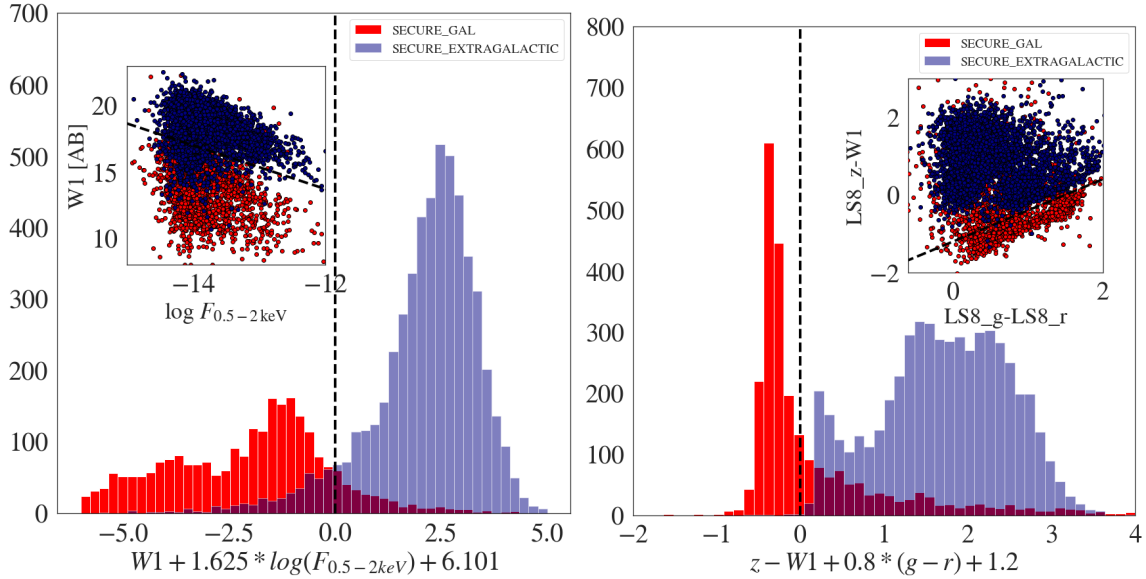
**Fig. 8.** Decision tree we adopted to assign each eFEDS point source to the Galactic or extragalactic class. First we classified the sources on the basis of the most secure methods (e.g. high-confidence redshift) and then proceeded with less reliable methods (e.g. based on colours) on the sources remaining in the pool, creating less pure samples. The numbers listed at each step include all sources, i.e., they also include those with an insecure counterpart ( $CTP\_quality < 2$ ).

and 2976 classified as Galactic (2566 secure and 410 likely). All these numbers are reported in Table 6.

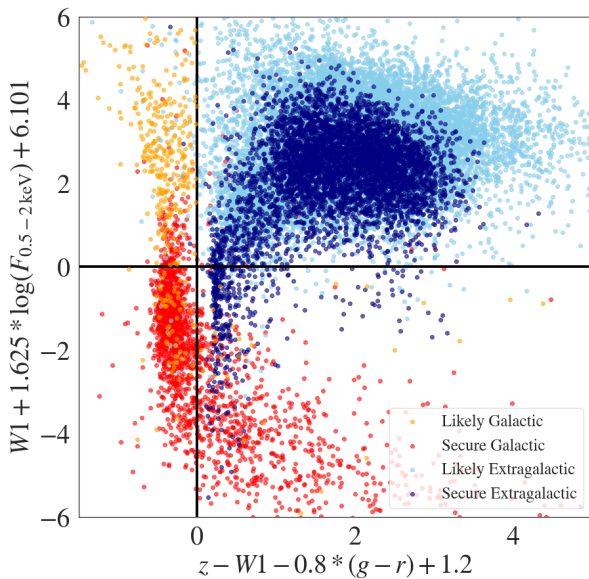
**6.2. Validation of the classification using external samples**

We carried out sanity checks of the classification framework against two external catalogues whose members are expected to

be almost completely extragalactic in nature: the Faint Images of the Radio Sky radio component catalogue (FIRST; White et al. 1997), and the Gaia-unWISE AGN candidate catalogue (GUA; Shu et al. 2019). Whilst the GUA sample is derived from underlying datasets similar to those used in our own source classification logic, the machine-learning methods used by Shu et al. (2019) make this at least a semi-independent test sample. Simple



**Fig. 9.** Distribution of sources flagged as “secure Galactic” (red) and “secure extragalactic” (blue) in the  $g-r$  vs  $z-W1$  (left) and  $W1$  vs. X-ray (right) planes that were used to determine a line separator (black line) to classify sources in steps 5, 6, and 8 of the flowchart presented in Fig. 8. The line separator on the right has fewer Galactic sources that fall into the extragalactic locus. However, with the line separator defined on the left, only a handful of extragalactic sources fall into the Galactic locus. This makes this classifier more efficient when the four photometric points are available.



**Fig. 10.** Four eFEDS X-ray source classes “secure Galactic” (red), “likely Galactic” (orange), “secure extragalactic” (blue), and “likely extragalactic” (cyan) defined in the flowchart presented in Fig. 8, distributed according to their distance from the two lines defined in Fig. 9. Three of the four wedges thus defined contain extragalactic or Galactic samples that are up to 99% pure (see text for details).

positional matches were made against our best matching optical counterpart positions, with a search radius of 3 arcsec for the FIRST radio component catalogue<sup>9</sup> and 1 arcsec for GUA (GUA objects were considered when they had  $\text{PROB\_RF} > 0.8$ ). We examined the rate at which sources we classified as Galactic or extragalactic (both secure and likely) were matched to objects in these external catalogues (see Table 6). There is a very low

<sup>9</sup> We only considered the radio components and made no attempt to handle complex sources appropriately.

**Table 6.** Comparison of our classification scheme against two (semi-) independent reference catalogues: the FIRST radio component catalogue (White et al. 1997), and the *Gaia*–unWISE AGN candidate catalogue (GUA, Shu et al. 2019).

| Ref. sample | Total matches | Extragalactic |        | Galactic |        |
|-------------|---------------|---------------|--------|----------|--------|
|             |               | Secure        | Likely | Likely   | Secure |
| All eFEDS   | 27 369        | 5377          | 19 016 | 410      | 2566   |
| FIRST       | 796           | 376           | 414    | 1        | 5      |
| GUA         | 6357          | 2924          | 3425   | 3        | 5      |

**Notes.** The low rate at which our classification logic classifies both radio sources and AGN candidates as secure Galactic or likely Galactic suggests that our classifications are robust.

rate of apparent disparities between our classifications and those that may be derived by matches to the external catalogues. For example, only 0.19% of secure Galactic sources have a radio counterpart in FIRST, compared to 7.0% of the secure extragalactic sample. Likewise, only 0.19% of the secure Galactic sources are matched to candidate AGN from Shu et al. (2019), compared to 54% of the secure extragalactic subsample.

### 6.3. Very nearby galaxies

Unlike what happens in pencil-beam surveys, there are numerous very nearby and thus resolved galaxies within eFEDS. Vulic et al. (2022) searched for eFEDS sources within the D25 ellipse of the sources in the Heraklion Extragalactic CATALOGUE (HECATE) of nearby galaxies (Kovlakas et al. 2021). For the 100 HECATE galaxies with a nearby eFEDS source, 93 out of 100 are consistent with the counterpart proposed here by the combination of NWAY ASTROMATCH. For the remaining 7 cases (ID\_SRC 7551, 12847, 2671, 22198, 17437, 29989, and 20952; see Fig. A.1), the counterparts identified in this work fall within

the HECATE galaxies, but do not coincide with the centre of the galaxy, but rather with a source that could be either an ULX in the galaxy or an extragalactic source in the background. For these 7 sources, dedicated studies will be needed to identify the exact origin of the X-ray emission.

## 7. Photometric redshifts

Photo- $z$  of AGN and X-ray selected sources in general have developed dramatically in the past ten years, bringing the redshift accuracy and the fraction of outliers (usual quantities measured to assess the quality of the photo- $z$ ) comparable to those measured for normal galaxies.

Regardless of whether photo- $z$  are computed via SED fitting or via machine-learning, accurate photo- $z$  for AGN are less straightforward to obtain than those for non-active galaxies (see Salvato et al. 2018, for a review of the topic), mainly because for each multi-wavelength data point, the relative contribution of host and nuclear emission is unknown and redshift dependent. Redshift, however, is the parameter that we are trying to determine. To add to the difficulty, the impact of dust extinction and variability must not be forgotten. Variability is an intrinsic property of AGN. Especially for wide-area surveys, where data are taken over many years, this can noticeably affect the accuracy of photo- $z$  if it is not accounted for (e.g. Simm et al. 2015), as was possible to do in COSMOS, for instance (Salvato et al. 2009, 2011; Marchesi et al. 2016). In eFEDS, we also have to face the issue that the photometry is not homogenised, and different surveys cover different parts of the field at different depth and there are different ways of computing the photometry (Kron, Petrosian, apertures, model, etc). In the following we describe the procedure we adopted to compute photo- $z$  using LePHARE (Arnouts et al. 1999; Ilbert et al. 2006). We then proceed with an estimate of the reliability of the photo- $z$  and a comparison with DNNZ, an independent computation of photo- $z$  using machine-learning (Nischizawa et al., in prep.).

### 7.1. Photo- $z$ computation

We computed the photo- $z$  for the sources classified as extragalactic. In order to minimise systematic effects, we used different types of photometry, depending on the survey; in particular, we tried to avoid photometry derived from models for the extended and nearby sources because usual models are good representation of point-like, disk-like, and bulge-like sources, but are unable to represent a local Seyfert galaxy, for example, in which nuclear and host components both contribute to the total flux. For this reason, we used total fluxes from GALEX; Kron and cmodel photometry from HSC, depending on whether the source was extended (see below); and GAAP (Gaussian Aperture and Photometry) from KiDS+VIKING. From VHS, we adopted Petrosian photometry as it appears to agree better with the VISTA/VIKING photometry. All the photometry was corrected for Galactic extinction using  $E(B-V)$  from LS8. Depending on whether the source was in the area covered by KiDS+VIKING, within HSC but outside KiDS and outside HSC, different bands were available<sup>10</sup>.

<sup>10</sup> In particular, for HSC, in the S19A release available to us at the time of this work, photometry in  $r2$  and  $i2$  filters is provided. However, the filters have changed during the survey, and depending on the coordinates of the sources, the fraction of data obtained with the original or the new filters changes. In order to account for this at any location, we adopted the filter that was used to obtain at least 50% of the data. This solution is not optimal and will affect the quality of the photo- $z$  in some areas.

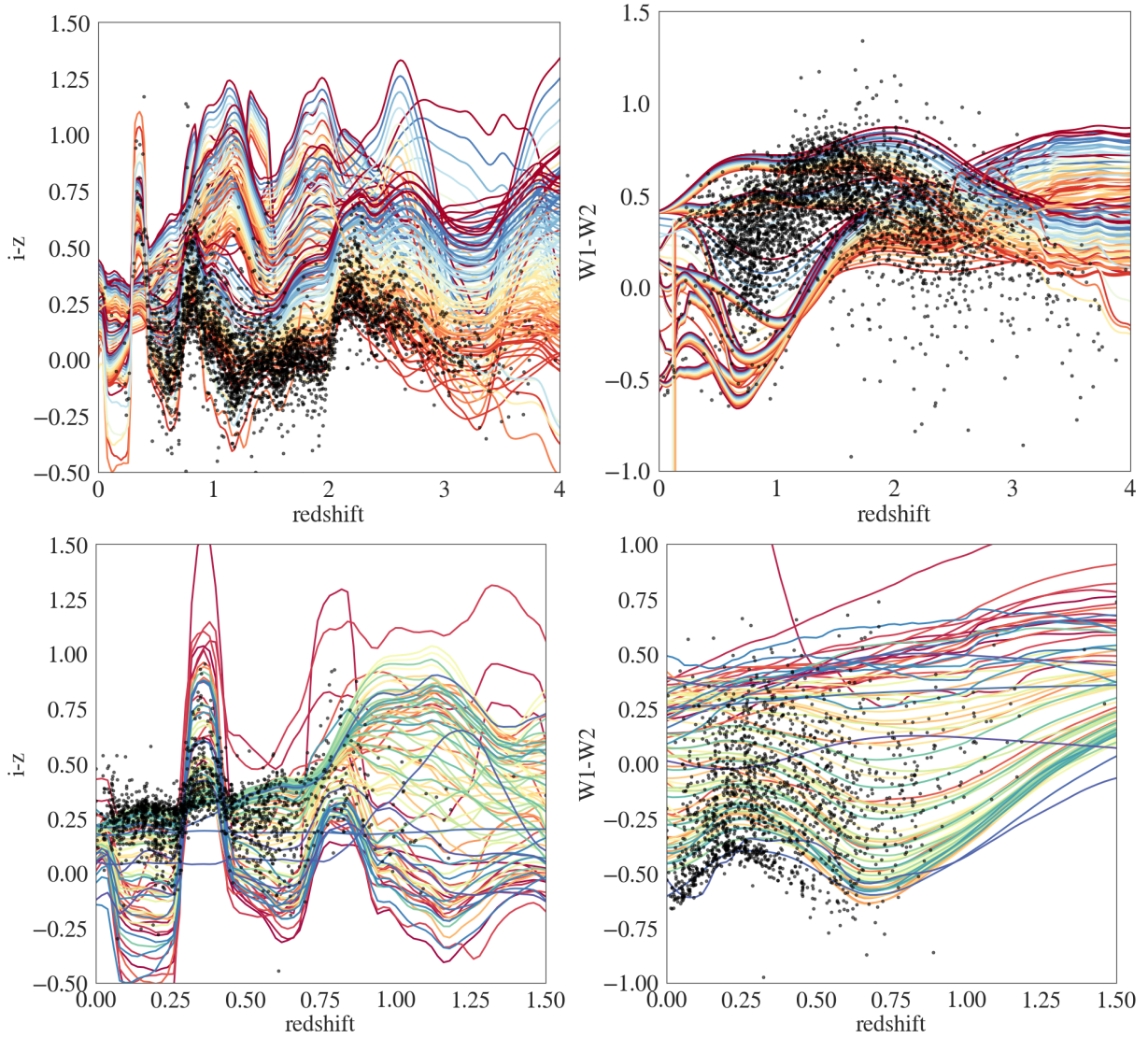
The computation of the photo- $z$  followed the procedure already outlined in previous works (Salvato et al. 2009, 2011; Fotopoulou et al. 2012; Hsu et al. 2014; Marchesi et al. 2016; Ananna et al. 2017), where sources were treated differently, depending on whether the optical images indicate them being extended (EXT) or a point-like/unresolved (PLIKE), following Sect. 6. This step is particularly important, as sources in the two samples are treated differently, using different priors and templates.

In addition, the fitting templates were selected on the basis of the X-ray depth and coverage of the surveys, keeping in mind that bright AGN, for instance, will mostly be absent in a deep pencil-beam survey. These surveys are characterised instead by host-galaxy dominated sources. Given the similar X-ray depth, the libraries used in Ananna et al. (2017) for the Stripe-82X survey were a good starting point for our work on eFEDS. However, a new library of templates for AGN and hybrids (AGN and host) was recently presented in Brown et al. (2019). The authors used photometry and archival spectroscopy of 41 AGN to create an additional set of 75 new hybrid templates. With respect to previous AGN templates, they have the advantage that they are empirical for the entire wavelength coverage and that the contribution from the host and AGN components is fully taken into account when the final SED is created, including dust attenuation and emission lines.

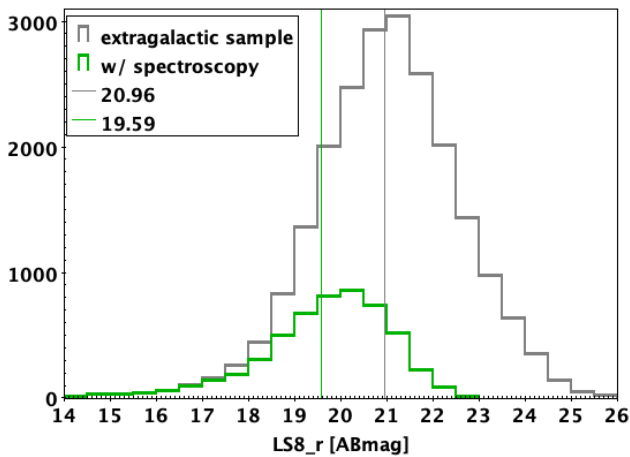
eFEDS is particularly rich in sources with reliable spectroscopy (see Sect. 2.3), allowing for a better tuning of the templates to be used to compute the photo- $z$ . To optimise the template choice, the colours of all the sources with reliable spectroscopy were plotted as a function of redshift, together with the theoretical colours from all the templates available (Fig. 11 illustrates this for  $i-z$  and  $W1-W2$  for the EXT and PLIKE samples, respectively, for all the templates that we ultimately adopted). In selecting the templates, we tried to limit their number (to control degeneracy in the redshift solution), while at the same time compiling a list representative of the entire population. The sources that in Fig. 11 are outside the parameter space covered by the templates can be interpreted in various ways, from problems in the photometry of the specific objects due to blending with neighbours or variability or lack of certain features in the templates. While we will further investigate this latter possibility for the future eROSITA surveys, here we recall that the figures are representative of only two colours, while in selecting the templates, we study all the colours that our photometric set allows.

An important point to keep in mind is the fact that despite being rich, the available spectroscopic sample is not representative of the entire eFEDS population, as is shown in Fig. 12. For this reason, the final library should also include some templates for types of sources that are expected to be present in eFEDS, but have not necessarily been identified so far. In particular, we created a set of templates using the archetype of type 1 AGN from the counterparts of ROSAT/2RXS (Salvato et al. 2019) observed within SDSS-IV/SPIDERS presented in Comparat et al. (2020), extended in UV and mid-infrared with various slopes. For the non-empirical templates, reddening was also considered, using the extinction law of Prevot (Prevot et al. 1984) with  $E(B-V)$  values from 0 to 0.4 in steps of 0.1. The selected templates are presented in Appendix B.

As output, LE PHARE provides the best value for the best photo- $z$  together with the upper and lower 1, 2, and 3 $\sigma$  error, the best combination of template, extinction law, and extinction value, the quality of the fit, and the pdz, the latter being the redshift probability distribution defined as  $pdz = \int F(z) dz$



**Fig. 11.**  $i-z$  and  $W1-W2$  colours of eFEDS extragalactic sources as a function of their (reliable) spectroscopic redshift (black points). The tracks of theoretical colours as a function of redshift derived from all the templates used in this work for the PLIKE (*upper panels*) and EXT (*lower panels*) samples are overlotted, as listed in Appendix B.

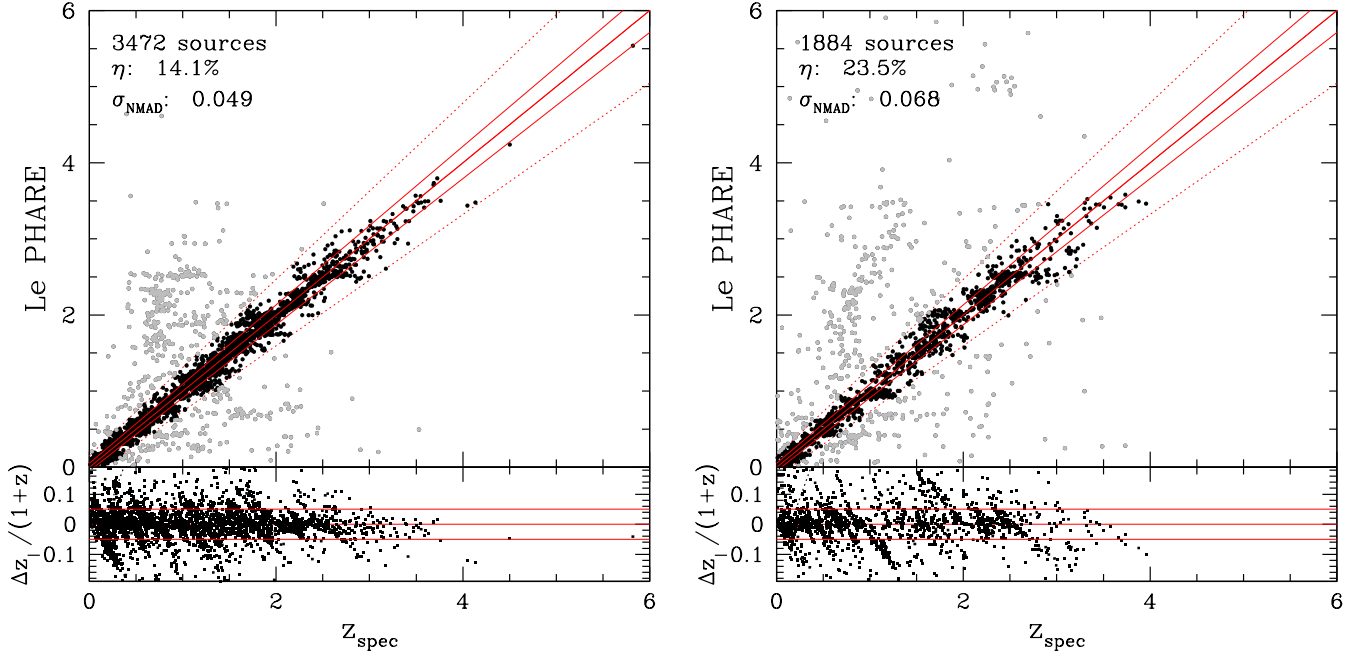


**Fig. 12.** Magnitude distribution for the entire extragalactic sample and the subsample for which reliable spectroscopy is available. The vertical lines indicate the mean values of the two samples.

between  $z_{\text{best}} \pm 0.1(1 + z_{\text{best}})$ , with  $z_{\text{best}}$  being the photo- $z$  value corresponding to the best fit. For photo- $z$  computed with a sufficiently high number of photometric bands covering the entire SED (e.g. XMM-COSMOS; Salvato et al. 2009), a high value of  $\text{pdz}$  can be safely translated into reliability of the photo- $z$ . As was discussed in Brescia et al. (2019), this is not the case when the photometric set is not rich, and  $\text{pdz}$  can be high also for a poor fit just because there are no sufficient constraints.

## 7.2. Reliability of photo- $z$

The final comparison between photo- $z$  and spec- $z$ , considering EXT and PLIKE sources together, for the area within KiDS+VIKING and within HSC, but outside KiDS+VIKING, is shown in Fig. 13. We used the standard metrics to measure the quality of photo- $z$  (see Salvato et al. 2018, for more details about definitions). (a) The fraction of outliers  $\eta$ : it highlights the fraction of sources with unexpectedly large errors, and it is defined as the fraction of sources for which  $|z_{\text{phot}} - z_{\text{spec}}| / (1 + z_{\text{spec}}) >$



**Fig. 13.** spec-z vs. photo-z from LE PHARE for the sources inside KiDS+VIKING (*left panel*) and inside HSC, but outside KiDS+VIKING (*right panel*). The sources in grey are considered outliers, and the red lines correspond to (1)  $z_{\text{phot}} = z_{\text{spec}}$  (thick solid), 9)  $z_{\text{phot}} = z_{\text{spec}} \pm 0.05 \times (1+z_{\text{spec}})$  (solid), and (3)  $z_{\text{phot}} = z_{\text{spec}} \pm 0.15(1+z_{\text{spec}})$  (dotted).

**Table 7.** Fraction of outliers and accuracy for LE PHARE and DNNz computed using the extragalactic sources with secure spectroscopic redshift, split by area.

| Area         | Sample | LE PHARE                         |        |                        | DNNz                             |        |                        |
|--------------|--------|----------------------------------|--------|------------------------|----------------------------------|--------|------------------------|
|              |        | $N_{\text{spec}}/N_{\text{tot}}$ | $\eta$ | $\sigma_{\text{NMAD}}$ | $N_{\text{spec}}/N_{\text{tot}}$ | $\eta$ | $\sigma_{\text{NMAD}}$ |
| Inside KiDS  | PLIKE  | 2331/6808                        | 17.8%  | 0.048                  | 2280/6808                        | 21.8%  | 0.045                  |
|              | EXT    | 1141/4598                        | 6.7%   | 0.054                  | 1139/4598                        | 6.3%   | 0.032                  |
|              | TOTAL* | 3472/11 406                      | 14.1%  | 0.049                  | 3419/11 406                      | 16.7%  | 0.039                  |
| Outside KiDS | PLIKE  | 1399/8532                        | 28.6%  | 0.081                  | 1325/8532                        | 34.9%  | 0.075                  |
|              | EXT    | 488/4221                         | 8.6%   | 0.040                  | 483/4221                         | 4.9%   | 0.026                  |
|              | TOTAL* | 1887/12 753                      | 23.8%  | 0.068                  | 1808/12 753                      | 27.2%  | 0.052                  |
| Outside HSC  | PLIKE  | 15/145                           | 42.9%  | 0.164                  | N/A                              | N/A    | N/A                    |
|              | EXT    | 3/89                             | 0.0%   | 0.039                  | N/A                              | N/A    | N/A                    |
|              | TOTAL  | 18/234                           | 33.3%  | 0.122                  | N/A                              | N/A    | N/A                    |

**Notes.** In each row, the difference between the numerators in the  $N_{\text{spec}}/N_{\text{tot}}$  columns provides the number of sources with spectroscopy for which DNNz could not provide a photo-z, mostly because the HSC photometry is saturated for these sources.

0.15 (e.g. Hildebrandt et al. 2010). (b) Accuracy  $\sigma_{\text{NMAD}}$ , which describes the expected scatter between predictions and truths and is defined as  $1.48 \times \text{median}(|z_{\text{phot}} - z_{\text{spec}}|/(1+z_{\text{spec}}))$  (Ilbert et al. 2006).

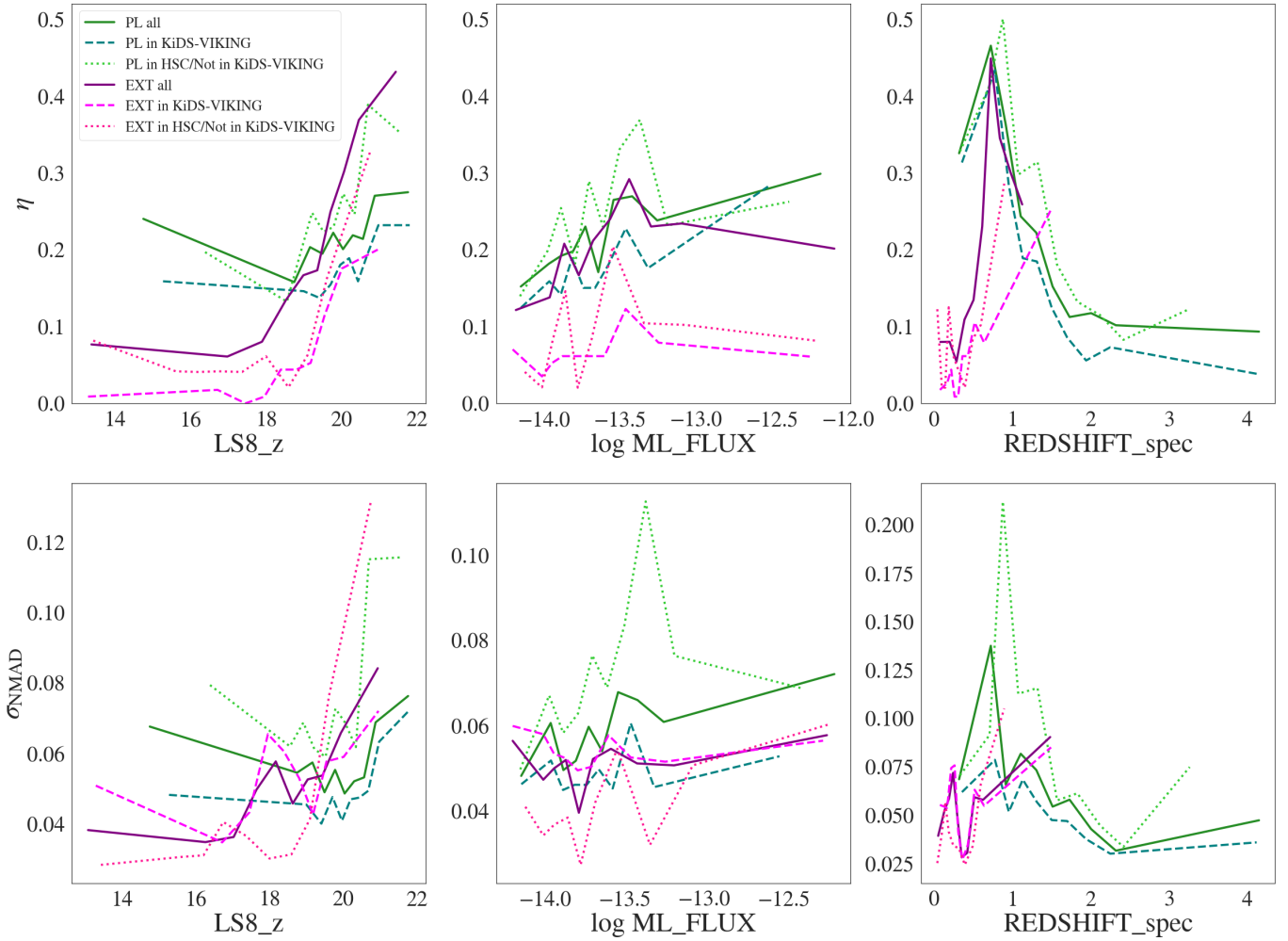
The results are listed in Table 7. Figure 14 shows the same results, but split as a function of  $z$ -band magnitude from LS8, X-ray flux, and spectroscopic redshift.

Ideally, for the best computation of photo-z, in particular, for sources dominated by emission lines such as AGN, photometry from broad-band filters across the entire spectral range should be complemented by narrow-band and near-infrared photometry and should be homogenised (e.g. Salvato et al. 2009, 2018). While narrow-band photometry is not available, at least some of the surveys provide homogenised photometry. For near-infrared photometry, the VISTA/VHS data are not sufficiently deep. The effect on the photo-z is clearly visible in all the panels of Fig. 14,

where the fraction of outliers is usually higher and the accuracy lower (high value of  $\sigma_{\text{NMAD}}$  in the area without VIKING coverage; dotted lines). Not only are the near-infrared data shallow outside the KiDS+VIKING area, they are also just a collection of photometric points computed in different ways, simply matched in coordinates. For this reason, based on the footprints shown in Fig. 1, we can think of the photo-z in eFEDS as divided into three regions that reflect the quality of the available photometry: the inner area is covered by deep forced photometry in KiDS+VIKING; the area that is within HSC, but outside KiDS+VIKING, for which some near-infrared information is provided by the shallow VISTA/VHS; and the area outside HSC for which the optical photometry is provided by LS8 alone.

The lack of deep near-infrared data also creates an unusual number of sources at high- $z$  ( $z > 3$ ), most of which are most likely incorrect. For example, the number of sources with





**Fig. 14.** Fraction of outliers (*top*) and accuracy (*bottom*) as a function of magnitude ( $z$  from LS8), 0.2–2.3 keV X-ray flux, and spectroscopic redshift split for type (EXT/PLIKE) and in area (with/without VIKING coverage). The  $x$ -axes are binned in equal numbers of elements, taking the quantiles between [0,1] in steps of 0.1. In other words, all lines also account for the size of each spectroscopic subsample.

photo- $z > 3$  is 188 within KiDS and 819 in the HSC area outside KiDS, although the area is about the same size. Within KiDS+VIKING, LE PHARE correctly estimates the redshift for 40 of the 55 (72.7%) sources that are spectroscopically confirmed to be at a redshift higher than 3. Most of these high- $z$  sources in excess can be easily identified and flagged by noticing that they are characterised by having high  $pdz$  even though they are in the area outside the HSC, that is, with a very limited number of photometric points to be fitted (see Sect. 7.4).

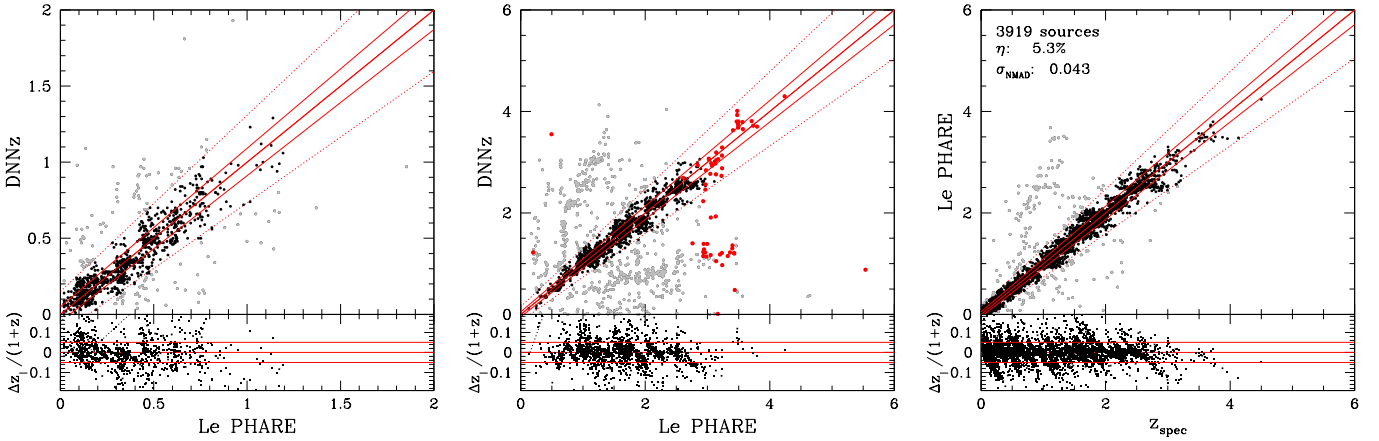
Figure 14 also shows how the accuracy degrades and the fraction of outliers increases for the PLIKE that are X-ray bright (top, second panels from the left). These sources are dominated by the AGN component with an SED close to a power law, for which the lack of narrow-band photometry that would identify the emission lines does not allow breaking the degeneracy in the redshift solutions. However, in eFEDS, there are only 47 extragalactic sources with an X-ray flux above  $5 \times 10^{-13}$  erg cm $^{-2}$  s $^{-1}$ , and a reliable spectroscopic redshift is available for 39 of them, so that the low quality of the photo- $z$  for these sources has only a limited effect. Finally, Fig. 14 shows an undesired high fraction of outliers at low redshift, where photo- $z$  values for normal galaxies are usually extremely accurate. The problem for AGN probably originates from the fact that both KiDS and HSC photometry are based on fitted models and not on total fluxes. Models are not able to account properly for the contribution of the nuclear

component that is comparable to the one from the host. Photometry from models can represent sources at very low redshift well, where the AGN contribution is negligible with respect to that from the host, and at high redshift, where the flux is dominated by the AGN component.

As already highlighted in the past, it is always easier to obtain a reliable photo- $z$  for galaxy-dominated sources with characteristic breaks in the SED. AGN-dominated sources are degenerate in the redshift solution, especially when little photometry is available, even within the KiDS+VIKING area (compare the dashed lines for EXT and PLIKE).

### 7.3. Comparison with DNNz

Within the HSC collaboration, the computation of photo- $z$  is available in many flavours. The method that performs better on AGN is DNNz (Nishizawa et al., in prep.). It is based on machine-learning and exclusively uses HSC photometry, trained on the rich spectroscopic sample available for both AGN and normal galaxies within the entire HSC region (beyond the area in common with eFEDS). The DNNz is based on the multi-layer perceptron (MLP) that takes the  $cmodel$  flux, PSF-matched aperture flux, and the second-order moment size measured at five HSC filter bands as inputs, and takes posterior probability as an output. In total,  $3 \times 5$  inputs and output PDF were binned in



**Fig. 15.** Direct comparison between photo- $z$  computed in this work with LE PHARE and DNNZ, within the HSC area for all the EXT (*left panel*) and PLIKE (*central panel*) sources. By construction, true EXT sources should not have spectroscopic redshift exceeding  $z \approx 1$ . It is not possible to decide a priori whether the photo- $z$  are incorrect or if the sources were placed erroneously in the EXT sample due to some issue of the photometry. In the *middle panel*, we highlight the sources that have spectroscopic redshift higher than 3 in red (see text for details). *Right panel*: Comparison between photo- $z$  from LE PHARE and spec- $z$  for the sources for which LE PHARE and DNNZ agree.

**Table 8.** Fraction of outliers for the photo- $z$  computed with LE PHARE (top) and DNNZ (bottom), split by area, with respect to the number of sources with spectroscopic redshift for which the two methods agree (first row) or disagree (second row), following the definition in Sect. 7.3.

| Outlier fraction           |                 |                 |
|----------------------------|-----------------|-----------------|
| LE PHARE                   | In KiDS         | Outside KiDS    |
| LE PHARE and DNNZ agree    | 3.7% [99/2663]  | 8.5% [107/1256] |
| LE PHARE and DNNZ disagree | 49.9% [377/756] | 51.3% [284/553] |
| DNNZ                       | In KiDS         | Outside KiDS    |
| LE PHARE and DNNZ agree    | 4.0% [106/2663] | 6.8% [85/1256]  |
| LE PHARE and DNNZ disagree | 56.1% [424/756] | 64.0% [354/553] |

**Notes.** The table clearly indicates that when LE PHARE and DNNZ disagree, DNNZ has a higher fraction of outliers among the spectroscopic sample, while when the two codes agree, the difference in fraction of outliers is marginal. To define agreement, we used  $1 + \text{mean}(\text{LEPHARE}, \text{DNNZ})$ . The small difference in the fraction of outliers for the two methods when they agree depends on how close they are to the real spectroscopic value.

100 bins from  $z = 0$  to  $z = 7$ . We have five hidden layers, and each layer has 100 nodes that are fully connected to the nodes in the neighbouring layers. With a 50k spectroscopic sample, it takes almost a whole day to train this machine with NVIDIA GeForce RTX 2080Ti GPU.

One interesting feature of DNNZ is that it was trained for any type of extragalactic source, without any particular tuning for AGN. In Table 7, the performances of DNNZ are directly compared with the output from LE PHARE. Remarkably, the accuracy of DNNZ is higher in general than for LE PHARE, although with a higher fraction of outliers.

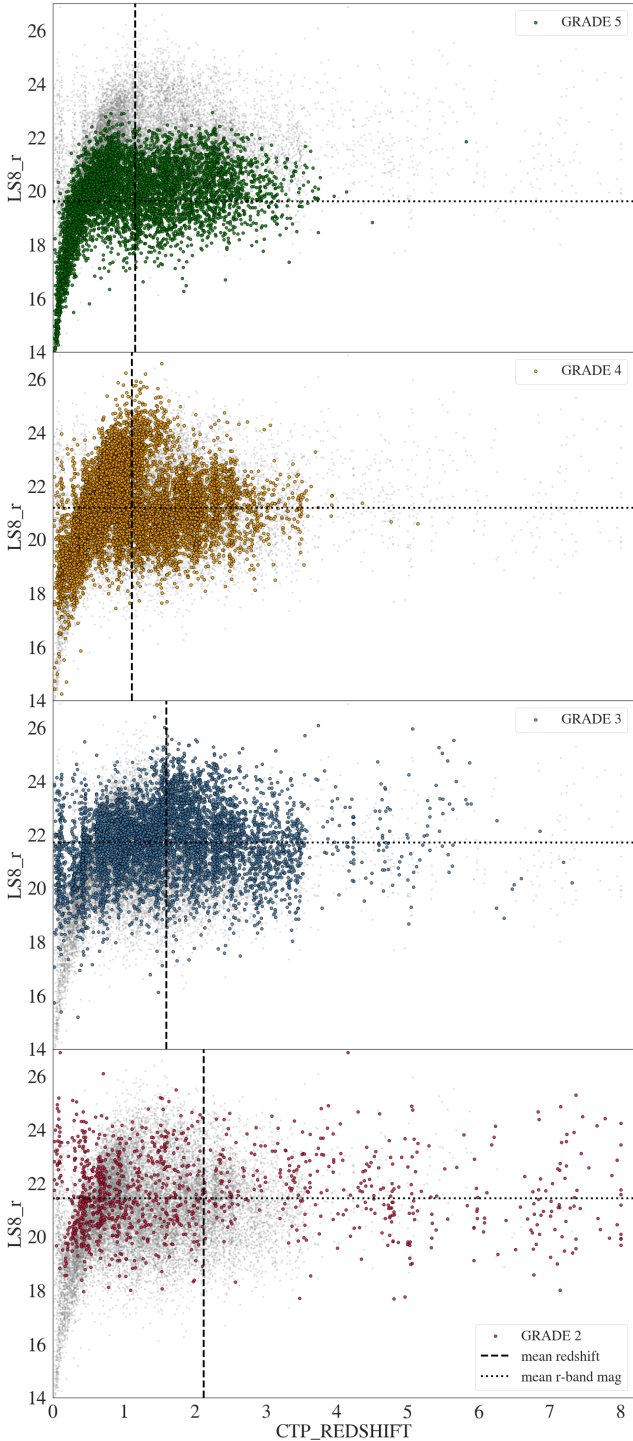
Interestingly, although only HSC photometry was used, DNNZ also shows a remarkable difference in the quality of the photo- $z$  for the sources within or outside the area covered by KiDS+VIKING. This is probably due to the combined photometry from the filters  $r$  and  $r2$  and  $i$  and  $i2$  that were changed during the survey. Most of the KiDS+VIKING area has been homogeneously observed only in  $i$  and  $r$  band, while the rest of the area has a mixture of observations. Taking this into account, we can compare LE PHARE and DNNZ in the area within KiDS+VIKING and split by TYPE. Figure 15 shows that both sets of photo- $z$  have some systematics (vertical and horizontal substructures) that are due on the one hand to the imbalance between galaxies and AGN in the training of DNNZ, and on the

other hand, to the degeneracies in the solution for power-law-dominated AGN and limited availability in photometry for LE PHARE.

However, when the photometry is sufficient and of good quality, SED fitting can correctly predict the redshift of AGN also when it is higher than 3 (middle panel of Fig. 15; sources in red). This is a current limitation for photo- $z$  computed via machine-learning because the sample of this type of source available for training is small (see Nishizawa et al., in prep.).

When the photo- $z$  from DNNZ is available, we can measure the mean photo- $z$  between the values proposed by the two methods for each source. Assuming this value is the right one, we have that for 60.6% of the extragalactic sources with  $\text{CTP\_quality} \geq 2$  DNNZ and LE PHARE agree ( $|z_{\text{LEPHARE}} - z_{\text{DNNZ}}| < 0.15 \times (1 + \text{mean}(z_{\text{LEPHARE}}, z_{\text{DNNZ}}))$ ). The comparison between LE PHARE and the spectroscopic redshift for 3919 sources with spec- $z$  is shown in the third panel of Fig. 15; the fraction of outliers with respect to the spectroscopic sample is extremely small and the accuracy is very high, comparable to the accuracy that is routinely obtained for normal galaxies, using purely broad-band photometry.

Table 8 summarises the result for DNNZ and LE PHARE separately, within and outside KiDS+VIKING. For the about 7500 sources for which the two methods provide differing



**Fig. 16.** eFEDS sources with reliable counterpart distributed in the magnitude vs redshift plane, split among the various `CTP_REDSHIFT_GRADE` classes. For each panel, we also indicate the mean value for redshift (vertical line) and magnitude (horizontal line). The entire population is shown in light grey.

results, the spectroscopic sample does not help to distinguish the best photo-z because the spectroscopic sample is very small (756 and 553 sources in the two areas, respectively) and not representative of the magnitude distribution in the sample (mean  $r$  value of the spectroscopic sample 20; the mean  $r$  value of the sample for which DNNZ and LE PHARE disagree is 21.5. See also the next section and Fig. 16). Photo-z derived via machine-learning

**Table 9.** Properties distribution for secure counterparts classified as extragalactic sources with `pdz` lower than a certain threshold.

| pdz threshold | N. sources | N. sources w/spec-z | N. sources w/spec-z & outliers | N. sources w/ $z_{\text{phot}} > 4$ |
|---------------|------------|---------------------|--------------------------------|-------------------------------------|
| <20           | 683        | 60/5287             | 46/60                          | 103/386                             |
| <30           | 1058       | 93/5287             | 67/93                          | 199/386                             |
| <40           | 1429       | 134/5287            | 94/134                         | 230/386                             |
| <50           | 1915       | 197/5287            | 121/197                        | 252/386                             |

**Notes.** The lower the `pdz`, the lower the quality of the fitting.

are well known to be very reliable only within the parameter space represented by the training sample and have little predictive power outside this space (e.g. Brescia et al. 2019). Keeping this in mind, we decided to rely on the prediction power of SED fitting and to rely on the results from LE PHARE. However, we also report the results from DNNZ and flag the sources for which LE PHARE and DNNZ agree or disagree (see Sect. 7.4).

#### 7.4. `CTP_REDSHIFT` and `CTP_REDSHIFT_GRADE` in the final catalogue

In the final catalogue we report the spectroscopic redshifts (regardless of their reliability) and the photo-z from both LE PHARE and DNNZ. In addition, for each source we summarise in the two columns `CTP_REDSHIFT` and `CTP_REDSHIFT_GRADE` our best knowledge of redshift and its reliability.

The column `CTP_REDSHIFT` lists original spectroscopic redshift when it is available and reliable (`NORMQ` = 3). The redshift is set to 0 for all the sources that are classified as GALACTIC (either SECURE or LIKELY) or for which no reliable redshift is available. To the remaining sources we assign the photo-z from LE PHARE.

Then in the column `CTP_REDSHIFT_GRADE` we provide a grade of confidence to the redshifts. The grades are listed below.

- `CTP_REDSHIFT_GRADE` = 5: this is a higher grade, assigned to the sources with reliable spectroscopic redshift. Of the 6591 sources in this category, 5377 are extragalactic sources and 1214 are Galactic (6465/6591 with `CTP_quality`  $\geq$  2).

- `CTP_REDSHIFT_GRADE` = 4: this is assigned to the sources for which the photo-z from LE PHARE and DNNZ agree (10 949 in total, 9643 of which have `CTP_quality`  $\geq$  2), because in the previous section we demonstrated that for this subsample, the fraction of outliers is very small and the accuracy very high. By construction, all the Galactic sources without spectroscopic redshift have `CTP_REDSHIFT_GRADE` = 4 because DNNZ and LE PHARE are set to zero and belong to this subsample (2995 sources).

- `CTP_REDSHIFT_GRADE` = 3: this is assigned to the sources for which LE PHARE and DNNZ disagree and `pdz` > 40 (6741 in total, 6057 of the sources with `CTP_quality`  $\geq$  2). The threshold at `pdz` > 40 was set by considering the fraction of outliers as a function of `pdz` in the sample with spectroscopic redshift (see Table 9). At the same time, we searched for the value of `pdz` that minimised the number of outliers and maximised the number of sources with  $z_{\text{phot}} > 4$ . The latter is suspiciously too high. This is due to the lack of deep photometry not only in the UV, but also in near-infrared: the large majority of these high- $z$  sources are concentrated in the area outside KiDS.

– CTP\_REDSHIFT\_GRADE = 2: this is assigned to the remaining sources for which LE PHARE and DNNZ disagree and  $\text{pdz} < 40$ , for which we are less confident about the photometric redshifts. This group includes only 1326 sources, 1092 of which have a CTP\_quality  $\geq 2$ .

Figure 16 shows the distribution of the sources for each of the REDSHIFT\_GRADE in the magnitude redshift plane. The mean value of the redshift and magnitude for each of the subsamples is also indicated.

### 7.5. Flagging sources likely associated with clusters of galaxies in the point-like sample

By construction, the eFEDS X-ray point-source catalogue is expected to be very little contaminated by clusters of galaxies; still, a low probability that a source is actually a cluster remains, as was shown in the simulations we performed for eFEDS (Liu et al. 2022c). Clusters end up in the point-like sample for many reasons (Willis et al. 2021). Most obviously, clusters with a small apparent size or at low detection likelihood can fall below the thresholds that are used to define the extension of the X-ray source. In addition, clusters could leak into the point-like sample because of source splitting and superimposition of a bright point source and a cluster.

With this in mind, we ran the multi-component matched filter cluster confirmation tool (MCMF; Klein et al. 2018, 2019) on the eFEDS point source catalog. We ran MCMF as in the eFEDS extended sources catalogue (Klein et al. 2022) after adjusting some of the parameters (e.g. limiting the area search from the X-ray position). As in Klein et al. (2022), we defined a “contamination fraction”,  $f_{\text{cont}}$ , which expresses the probability for an optical concentration of red galaxies to be a chance alignment along the line of sight to the X-ray source. This is the key selection criterion for selecting cluster candidates, and it immediately provides an estimate of the catalogue contamination. A catalogue created by selecting  $f_{\text{cont}} < a$  is expected to have a contamination fraction of  $a$ , assuming the input catalogue is highly contaminated.

Because of the high number density of sources in the point-like sample and/or the possibility that the emission from an actual cluster is split into many point sources, it can happen that many close X-ray sources point to the same optical cluster. A simple cut in  $f_{\text{cont}}$  will therefore yield a much larger sample of sources than real clusters in that catalog, causing the contamination fraction to be much higher than expected. To compensate for this, for each eFEDS point source that is close to an optical overdensity, an environmental flag is set to true for the source that is closest to the overdensity and that is at least 0.75 Mpc away from a cluster detected in the extent-selected sample (Liu et al. 2022a,c; Klein et al. 2022) at similar redshift. Only when the flag is set to true is the point-source further considered as a candidate for being a cluster.

For this latter subgroup of sources, following Klein et al. (2022), MCMF assigns a redshift to the cluster (via the red sequence). In addition, the photo- $z$  of the counterpart to the point-like sources is recomputed assuming they are passive galaxies. The two redshifts are then compared with the redshift computed by LEPHARE, as described in the previous section (Sect. 7).

Combining all the information described above, we define a new flag, Cluster\_Class, which indicates the possibility that an eFEDS X-ray (point-like) source is actually a cluster or belongs to a cluster.

– Cluster\_class = 5: CTP\_QUALITY  $\leq 1$  &  $f_{\text{cont}} < 0.2$  and the environmental flag set to true: the counterpart NWAY or ASTROMATCH is considered unreliable and the X-ray emission is more likely associated with a cluster (top left panel of Fig. 17; 120 cases).

– Cluster\_class = 4: CTP\_QUALITY  $\geq 2$  &  $f_{\text{cont}} < 0.2$  with the environmental flag set to true, the optical colours of the sources are typical of passive galaxies, and the redshift computed with LEPHARE coincides with the redshift of the optical cluster: the counterpart is reliable, and the point source is a galaxy member (possibly the BCG) of the optically detected cluster (top right panel in Fig. 17; 63 cases).

– Cluster\_class = 3: CTP\_QUALITY  $\geq 2$  &  $f_{\text{cont}} < 0.2$  and the environmental flag set to true and the redshift computed assuming an AGN template is consistent with the redshift of the optical cluster, but the optical colours of the counterpart are not typical of a passive galaxy: the counterpart is correct and the source is a cluster member (bottom left panel from the left of Fig. 17; 96 cases).

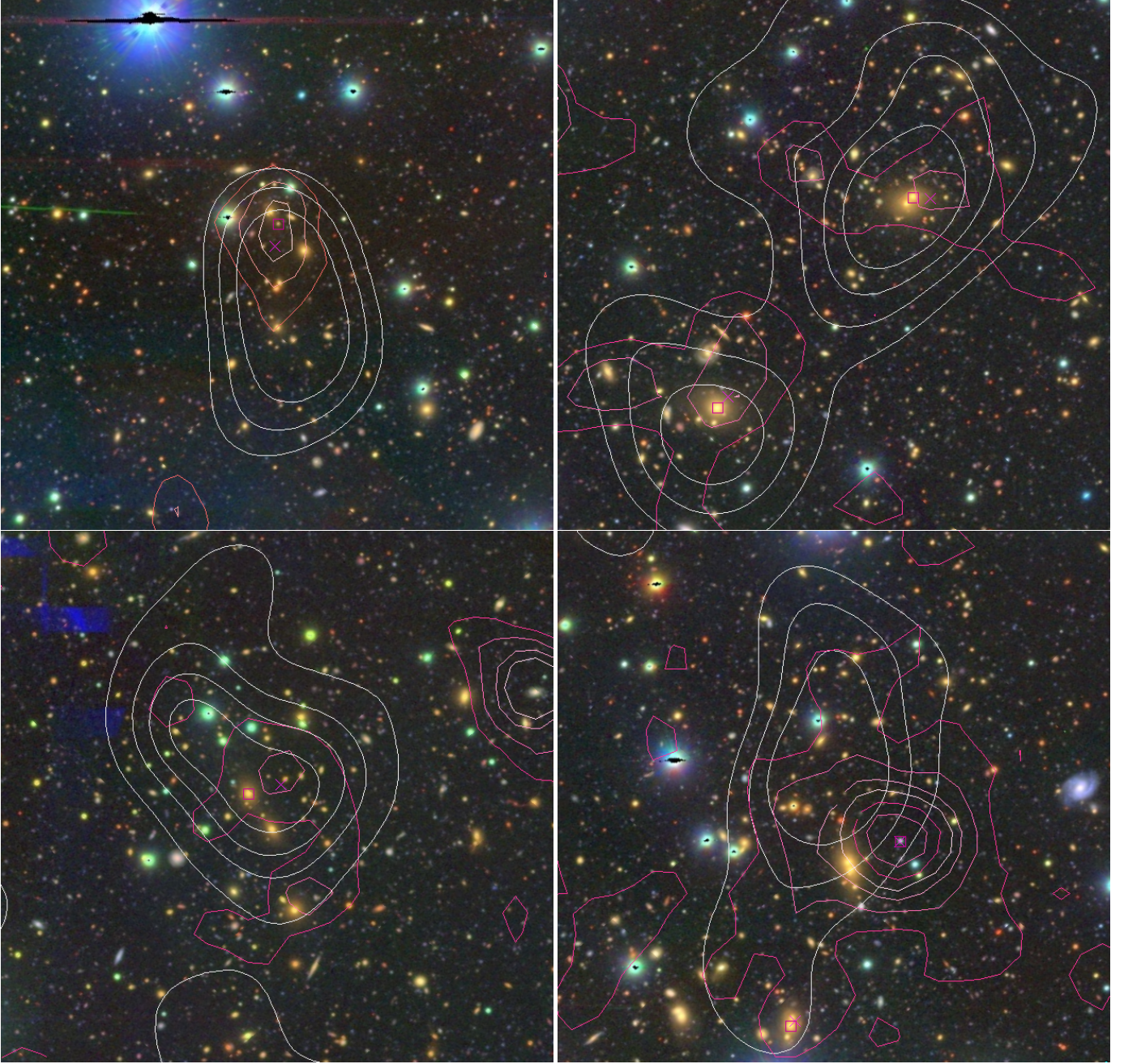
– Cluster\_class = 2: CTP\_QUALITY  $\geq 2$  &  $f_{\text{cont}} < 0.01$  and the environmental flag set to true, while the photo- $z$  computed by the three methods disagree: the counterpart is reliable, and the source (AGN) is just projected on a likely cluster (bottom right panel of Fig. 17; 67 cases).

In all the other cases, the X-ray emission is from a genuine point-source and likely not from the extended, hot intercluster medium. A dedicated effort is currently ongoing to confirm the secure clusters in the point-source catalogue and to characterise and measure the X-ray and radio properties of the confirmed clusters (e.g. Bulbul et al. 2022).

## 8. Data release

The catalogs listing the properties of the counterparts to eFEDS point-like sources in the main and hard samples samples (Brunner et al. 2022) associated with this paper are available via CDS/Vizier and via the web page at MPE dedicated to the eROSITA data release<sup>11</sup>. The list of the columns and their description for the two samples is available in Appendix D. Only the basic X-ray properties are listed here (columns 1-9). For the complete list, we refer to the catalogs released by Brunner et al. (2022). After the columns reporting the key X-ray properties of the sources, Cols. 10–36 report the results of the counterpart (CTP) association, followed by the key parameters from NWAY, ASTROMATCH, and HamStar. Next (Cols. 36–49) we present the photometry from the recent *Gaia* EDR3 release in the original photometric system, followed by all the collected photometry, corrected for extinction (Cols. 51–108). We recall that the HSC photometry from S19A in  $i$  and  $r$  bands was split into  $i$ ,  $i2$ , and  $r$ ,  $r2$ , and that Kron is listed for EXT sources, while cmodel is listed for PLIKE (see Sect. 7). Columns 109–117 list basic properties of the sources, such as whether they are within KiDS or HSC, while Cols. 118–126 list all the information related to spectroscopy when available. The output parameters from LEPHARE and DNNZ are listed in Cols. 127–148. The columns CTP\_REDSHIFT and CTP\_REDSHIFT\_GRADE summarise the redshift properties of the sources, as discussed in Sect. 4.2, and the column CLUSTER\_CLASS refers to the results presented in Sect. 7.5.

<sup>11</sup> <https://erosita.mpe.mpg.de/edr/eROSITAobservations/Catalogues/>



**Fig. 17.** *Top left, clock-wise:* four examples of `Cluster_Class` = 5,4,3,2, as described in Sect. 7.5. X-ray contours are plotted in white. The magenta cross indicates the X-ray position, and the magenta square indicates the counterpart selected in this paper. Magenta contours indicate the red-sequence galaxy density. The HSC  $g, r, i$  cutouts are  $5.5' \times 5.5'$  in size.

In addition to the catalogs, we provide the redshift distribution function and SED fitting of each source in the catalogue at direct request to the first author. An example is shown in Fig. C.1.

## 9. Discussion

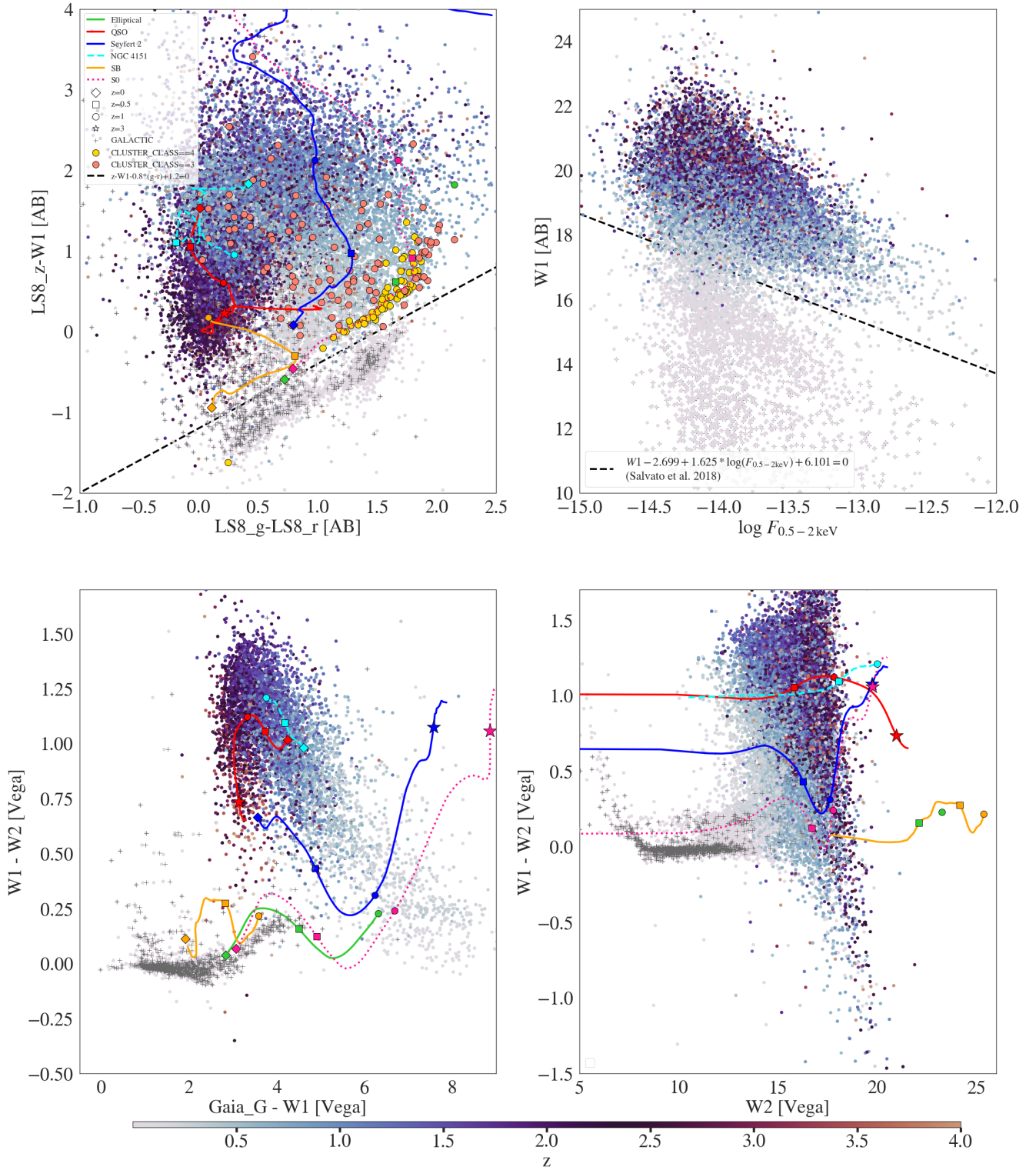
The size and depth of the eFEDS X-ray survey, combined with ancillary data both in photometry and spectroscopy, allows us to paint a comprehensive picture of the average population of X-ray sources that contribute the bulk of the cosmic X-ray background (CXB) flux at energies  $<10$  keV (see e.g. Gilli et al. 2007) in its Galactic and extragalactic content. The identification of the optical/IR counterparts, to a high degree of completeness

and reliability, as discussed here, will facilitate detailed population studies of X-ray active stars, Galactic compact objects, and AGN. Here we briefly outline the main properties of our sample by examining the distributions of the X-ray sources in various colour/redshift spaces in detail.

### 9.1. Population studies

Figure 18 shows the distribution of all the eFEDS sources with a secure counterpart (`CTP_quality`  $\geq 2$ ; see Sect. 5) in four different multi-band photometric spaces, chosen for their wide applicability to large areas of the sky.

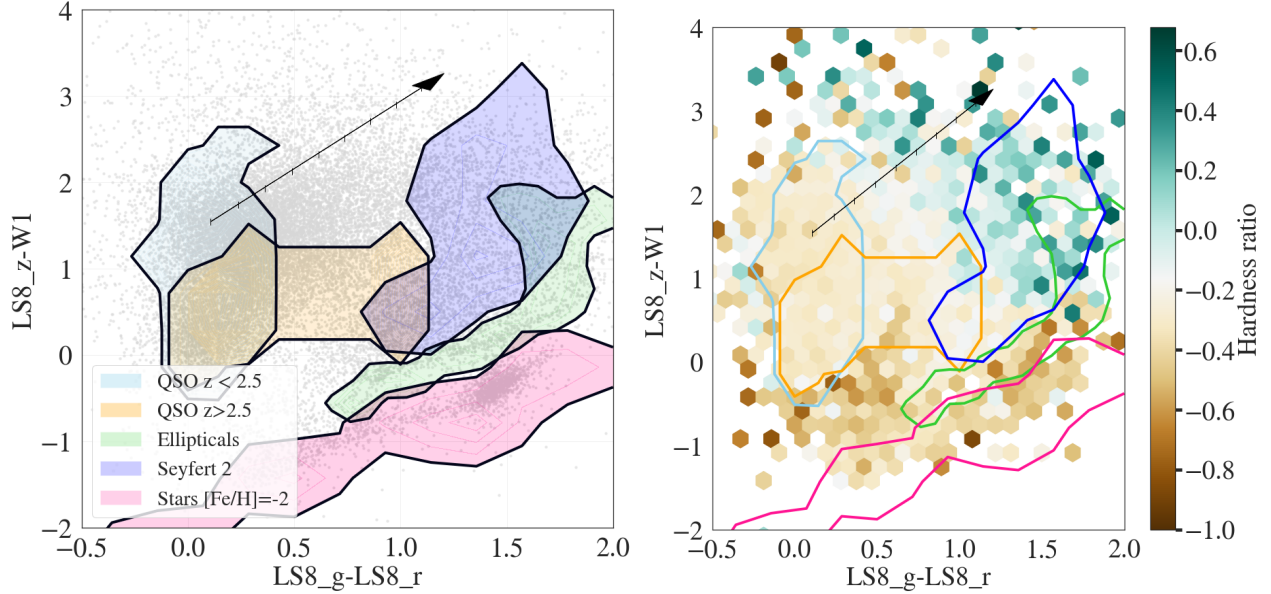
The top panel shows sources in the  $z$ - $W1$  versus  $g-r$  space, colour-coded by their redshift. A few representative tracks of



**Fig. 18.** Distribution of Galactic and extragalactic sources in eFEDS, colour-coded by redshift in the parameter spaces defined by typical colours and fluxes. For the plots with optical and mid-infrared colours, the track of templates characteristic of the population, which we also used to compute the photo- $z$ , are overlotted. The legend in the top left panel provides all the details for the four panels.

various classes of extragalactic objects are overlaid. In addition to the clear separation between Galactic and extragalactic objects we discussed in Sect. 6, the X-ray points identify clear sequences of unobscured QSO, obscured Seyferts, and inactive galaxies. The inactive galaxies are best represented by the S0 and elliptical tracks, suggesting that some of them are the sources that are

associated (or confused) with a cluster. The sources indicated by a yellow circle in the top left panel have CLUSTER\_CLASS = 3,4 indicating that they belong to a cluster. In most cases, they are the BCG (see Sect. 7.5). These sources are best fit by the template of a passive galaxy, as the spectra for those available also suggest (e.g. lack of emission lines from star formation, strong



**Fig. 19.** *Left:* classes of X-ray emitting sources are shown in colour-colour space. The clouds for QSOs, elliptical galaxies, and Seyfert 2 galaxies are derived from theoretical colour-redshift tracks computed with LePhare and a subset of SED templates used to determine photo- $z$  in this work. Seyfert 2 and elliptical tracks were limited to  $z < 1$ , while the QSO cloud is divided into  $z < 2.5$  and  $2.5 < z < 4$ . The black arrow indicates the evolution of a theoretical QSO at  $z=0.2$  in colour-colour space with increasing extinction. The ticks along this arrow are steps of  $\Delta_{E(B-V)}=0.1$ . The stellar cloud was derived from a 2D histogram of MIST/MESA isochrones assuming  $[\text{Fe}/\text{H}]=-2$  (Choi et al. 2016; Dotter 2016; Paxton et al. 2018). *Right:* same classes on top of the binned eFEDS data (CTP\_quality  $\geq 2$ ), colour-coded according to the average hardness ratio (see text for details).

HK lines). However, some of the spectra together with the clear features from a non-star-forming galaxy also reveal the presence of broad emission lines typical of AGN (see Bulbul et al. 2022).

The top right panel of Fig. 18 shows the distribution of points in the mid-infrared (WISE  $W1$ ) versus soft X-ray (0.5–2 keV) plane (same as Fig. 9), originally introduced in Salvato et al. (2018). X-ray bright objects above the dashed line are typically AGN, while most of the IR bright objects below the line are Galactic X-ray emitting stars, with some contamination from nearby extragalactic objects. These sources are rare, but given the size of eFEDS, their number is non-negligible. Thus, when using this plot for other surveys, the size of the survey must be accounted for. The larger the surveys, the less efficient the line separator.

The bottom right panel shows the distribution of the sources in the Wise-only  $W1-W2$  versus  $W2$  colour-magnitude plane<sup>12</sup>. This is widely used to classify point sources, as it easily separates stars, with  $W1-W2 \approx 0$ , from QSOs, with  $W1-W2 > 0.5$  (see e.g. Wright et al. 2010; Assef et al. 2013). Once more, the eFEDS X-ray selection reveals the full extent of the extragalactic (AGN) population with intermediate IR colours between AGN- and host-galaxy dominated, typical of either obscured (Seyfert 2) or low-luminosity AGN (e.g. Merloni 2016; Hickox & Alexander 2018).

Finally, the bottom left panel shows the distribution of the eFEDS sources in the optical/mid-infrared diagram defined by the “all-sky available”  $G-W1$  versus  $W1-W2$ , which is frequently used to separate QSO from stars in the *Gaia* catalog. As already pointed out in Sect. 6, 10% of the Galactic sources are too faint to be detected by *Gaia*. This is even more true for the extragalactic sources: the plot shows only 57% of the entire

eFEDS sample. However, the plot shows insights into the population that the first eROSITA All-Sky Survey (eRASS1) will uncover. As expected, the X-ray selected eFEDS sources contain beyond stars and (unobscured) QSOs a tail at high  $G-W1$  (i.e. bright mid-infrared, faint optical magnitudes) typical of inactive galaxies and/or mildly obscured AGN.

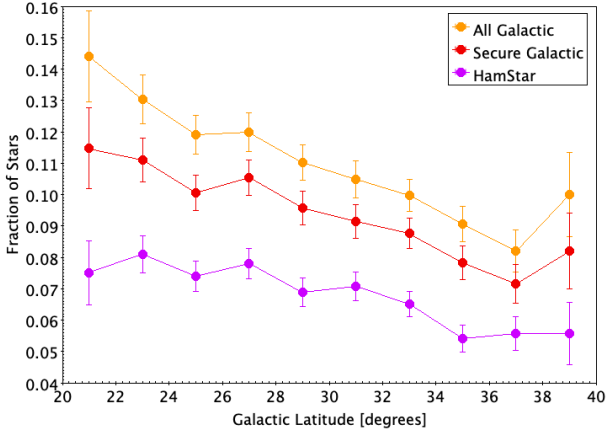
This is indeed confirmed by comparing the location of the extragalactic eFEDS sources in the  $grzW1$  plane with the X-ray hardness ratio measured from the X-ray counts in the bands in which eROSITA is most sensitive. The right panel of Fig. 19 shows the distribution of the sources in this plane, colour-coded by their average hardness ratio (defined as  $(H-S)/(H+S)$ , where  $H$  and  $S$  are the counts in the ranges 1.0–2.0 keV and 0.2–1.0 keV<sup>13</sup>, respectively), while the left panel highlights the loci of the most common classes of sources based on the distribution of template tracks. The hardest sources in the eROSITA band populate the optical/mid-infrared colour-space of Seyfert 2 galaxies and/or reddened QSOs. A detailed discussion of the X-ray spectral properties of the AGN in the main eFEDS sample will be presented in Liu et al. (2022b).

## 9.2. eFEDS stellar content

The eFEDS field spans a wide range of Galactic latitudes (from about +20 to about +40). Reassuringly, the fraction of the X-ray sources that are classified as Galactic (see Sect. 3 for the definition) increases towards low Galactic latitude, as shown in Fig. 20. The figure also confirms that the priors adopted by HamStar are not representative of all Galactic sources in

<sup>12</sup> For this plot, we use the Vega System, so that the user can compare the figure with similar ones prepared using AllWISE all sky.

<sup>13</sup> The hardness ratio is calculated using the columns ML\_CTS\_b1, ML\_CTS\_b2 and ML\_CTS\_b3 in Brunner et al. 2021 catalog:  $(\text{ML\_CTS\_b3} - (\text{ML\_CTS\_b1} + \text{ML\_CTS\_b2})) / (\text{ML\_CTS\_b3} + (\text{ML\_CTS\_b1} + \text{ML\_CTS\_b2}))$ .



**Fig. 20.** Fraction of all X-ray sources with  $\text{CTP\_quality} \geq 2$  that are classified as Galactic as a function of Galactic latitude. Purple symbols show objects classified with the HamStar method; red symbols show all “secure Galactic” objects, and orange symbols represent “Secure” and “Likely” Galactic objects (see Sect. 3 for the definition).

eFEDS. About 22.3% of the Galactic sources identified by NWAY/ASTROMATCH are fainter than the 19th magnitude (10% are not detected by *Gaia*).

## 10. Conclusions

We have presented the identification of the counterparts to the point sources in eFEDS listed in the main and hard catalogues (Brunner et al. 2022), together with the study of their multi-wavelength properties. eFEDS has a limiting flux of  $F_{0.5-2\text{keV}} \sim 6.5 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$  and is a factor of  $\sim 50\%$  deeper than the final eROSITA all-sky survey. It can therefore also be used as a forecast for eRASS:8, not only for the population that eRASS:8 will reveal, but also for the challenges that are ahead of us with respect to counterpart identification and redshift determination.

– Counterpart identification: We used NWAY (Salvato et al. 2019) and ASTROMATCH. In addition to spatial information from eFEDS and LS8, these codes use a prior based on the properties of a training sample of 3XMM sources that was tested on a validation sample of *Chandra* sources, made eFEDS-like in terms of positional accuracy. Each method has identified its own priors in a different way. For the validation sample, NWAY correctly identified 95% of the sources; only 2% of the sources have a possible second counterpart, compared with 89% and 10% for ASTROMATCH, but at the threshold adopted for  $p_{\text{any}}$  and  $\text{LR\_BEST}$ , both methods have very high completeness and purity (above 95%). These remarkable results, well above the predicted completeness and purity mentioned in Merloni et al. (2012), are due to three important factors: the development of new methods for identifying the correct counterparts, large samples of X-ray detected sources with known counterparts, and the availability of sufficiently deep, homogenised, multi-wavelength photometry from optical to mid-infrared over very wide areas from which the SED of these sources that were to be used as training was constructed. In the next two years, by the time eROSITA will have completed the final all-sky survey, the methods will continue to improve and the training/validation samples will increase in size. Most importantly, the coverage of the multi-wavelength catalogs that are used to identify the counterparts will be larger. While the DESI Legacy Imaging Survey DR9 (LS9; Dey et al. 2019) just became publicly available, the work on DR10 has started.

The survey will cover virtually all of the eROSITA-DE area of the sky at sufficient depth, which is possible because the DECAM data taken via the DeROSITAS survey are included (PI A. Zenteno). We predict that the identification of the counterparts for the entire eRASS will be of at least the same quality as eFEDS, also in the Galactic plane because the recently released *Gaia* EDR3 is included.

– Redshift determination: Given the lack of sufficiently deep near-infrared data outside the DES area (Sevilla-Noarbe et al. 2021), the possibility of obtaining reliable photometric redshifts via SED fitting will be low, at least until data from SpherEx (Doré et al. 2018) will be made available (launch planned for Summer 2024). However, as demonstrated in Nishizawa et al. (in prep.) and Borisov et al. (2021), the increasing size and completeness of the spectroscopic sample that can be used for the training will enable reliable photometric redshifts for any type of X-ray extragalactic source. For example, the spectroscopic follow-up of the eROSITA-DE sources planned via Vista/4MOST and SDSS-V/BHM will allow us to obtain redshifts for 80% of the sources detected by eRASS:3, thus limiting the need of photo-z, and at the same time, ensuring a high quality of photo-z that will use these spectroscopically confirmed sources as training.

*Acknowledgements.* We are thankful to the anonymous referee that with a careful reading of the paper helped us improving the manuscript. M.S. thanks Olivier Ilbert for the feedback provided on the computation of the photo-z. J.W. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC-2094 -390783311. M.K. acknowledges support by DFG grant KR 3338/4-1. B.M. acknowledges funding from European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 860744 (BiD4BEST). P.C.S. acknowledges support by DLR grants 50 OR 1901, 50 OR 2102. W.N.B. acknowledges financial support from NASA grant 80NSSC19K0961 and the V.M. Willaman Endowment at Penn State. This work is based on data from eROSITA, the soft X-ray instrument aboard SRG, a joint Russian-German science mission supported by the Russian Space Agency (Roskosmos), in the interests of the Russian Academy of Sciences represented by its Space Research Institute (IKI), and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The SRG spacecraft was built by Lavochkin Association (NPOL) and its subcontractors, and is operated by NPOL with support from the Max Planck Institute for Extraterrestrial Physics (MPE). The development and construction of the eROSITA X-ray instrument was led by MPE, with contributions from the Dr. Karl Remeis Observatory Bamberg and ECAP (FAU Erlangen-Nuernberg), the University of Hamburg Observatory, the Leibniz Institute for Astrophysics Potsdam (AIP), and the Institute for Astronomy and Astrophysics of the University of Tübingen, with the support of DLR and the Max Planck Society. The Argelander Institute for Astronomy of the University of Bonn and the Ludwig Maximilians Universität Munich also participated in the science preparation for eROSITA. We have made use of TOPCAT and STILTS (Taylor 2005, 2006) Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples). The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University. This paper is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center (ADC) at National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), National Astronomical Observatory of Japan. The Legacy Surveys consist of three individual and complementary



projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID 2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID 2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall  $z$ -band Legacy Survey (MzLS; Prop. ID 2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIR-Lab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Dúag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the US Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is [www.sdss.org](http://www.sdss.org). SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard and Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

## References

- Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, 259, 35
- Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, *ApJS*, 249, 3
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, *PASJ*, 70, S4
- Aihara, H., Armstrong, R., Bickerton, S., et al. 2018b, *PASJ*, 70, S8
- Aihara, H., AIsayyad, Y., Ando, M., et al. 2019, *PASJ*, 71, 114
- Ananna, T. T., Salvato, M., LaMassa, S., et al. 2017, *ApJ*, 850, 66
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
- Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, *ApJ*, 772, 26
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123
- Baldry, I. K., Liske, J., Brown, M. J. I., et al. 2018, *MNRAS*, 474, 3875
- Bianchi, L. 2014, *Ap&SS*, 354, 103
- Blanton, M. R., Bershad, M. A., Abolfathi, B., et al. 2017, *AJ*, 154, 28
- Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, *A&A*, 588, A103
- Boller, T., Schmitt, J. H. M. M., Buchner, J., et al. 2022, *A&A*, 661, A8 (eROSITA EDR SI)
- Borisov, V., Meshcheryakov, A., Gerasimov, S., & RU eROSITA catalog group 2021, ASP Conf. Ser., submitted [arXiv:2107.01891]
- Brescia, M., Salvato, M., Cavuoti, S., et al. 2019, *MNRAS*, 489, 663
- Brown, M. J. I., Duncan, K. J., Landt, H., et al. 2019, *MNRAS*, 489, 3351
- Brunner, H., Liu, T., Lamer, G., et al. 2022, *A&A*, 661, A1 (eROSITA EDR SI)
- Brusa, M., Comastri, A., Daddi, E., et al. 2005, *A&A*, 432, 69
- Brusa, M., Zamorani, G., Comastri, A., et al. 2007, *ApJS*, 172, 353
- Brusa, M., Urrutia, T., Toba, Y., et al. 2022, *A&A*, 661, A9 (eROSITA EDR SI)
- Buchner, J., Boller, T., Bogenberger, D., et al. 2022, *A&A*, 661, A18 (eROSITA EDR SI)
- Bulbul, E., Liu, A., Pasini, T., et al. 2022, *A&A*, 661, A10 (eROSITA EDR SI)
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102
- Civano, F., Elvis, M., Brusa, M., et al. 2012, *ApJS*, 201, 30
- Comparat, J., Merloni, A., Dwelly, T., et al. 2020, *A&A*, 636, A97
- Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, *MNRAS*, 392, 19
- de Vaucouleurs, G., de Vaucouleurs, A., Corwin, Herold G., J., et al. 1991, *Third Reference Catalogue of Bright Galaxies* (Berlin: Springer)
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168
- Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, ArXiv e-prints [arXiv:1805.05489]
- Dotter, A. 2016, *ApJS*, 222, 8
- Drinkwater, M. J., Byrne, Z. J., Blake, C., et al. 2018, *MNRAS*, 474, 4151
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astron. Geophys.*, 50, 5.12
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32
- Fotopoulou, S., Salvato, M., Hasinger, G., et al. 2012, *ApJS*, 198, 1
- Gaia Collaboration (Brown, A. G. A., et al.) 2018, *A&A*, 616, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2020, *A&A*, 649, A1
- Gilli, R., Comastri, A., & Hasinger, G. 2007, *A&A*, 463, 79
- Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, *AJ*, 131, 2332
- Hickox, R. C. & Alexander, D. M. 2018, *ARA&A*, 56, 625
- Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, *A&A*, 523, A31
- Hoffleit, D. 1964, *Catalogue of Bright Stars* (Cambridge: Harvard College Observatory)
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, *A&A*, 355, L27
- Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, *ApJ*, 796, 60
- Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *ApJS*, 199, 26
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841
- Jones, D. H., Read, M. A., Saunders, W., et al. 2009, *MNRAS*, 399, 683
- Klein, M., Mohr, J. J., Desai, S., et al. 2018, *MNRAS*, 474, 3324
- Klein, M., Grandis, S., Mohr, J. J., et al. 2019, *MNRAS*, 488, 739
- Klein, M., Oguri, M., Mohr, J. J., et al. 2022, *A&A*, 661, A4 (eROSITA EDR SI)
- Kovlakas, K., Zezas, A., Andrews, J. J., et al. 2021, *MNRAS*, 506, 1896
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Lang, D. 2014, *AJ*, 147, 108
- Liu, A., Bulbul, E., Ghirardini V., et al. 2022a, *A&A*, 661, A2 (eROSITA EDR SI)
- Liu, T., Buchner, J., Nandra, K., et al. 2022b, 2022b, *A&A*, 661, A5 (eROSITA EDR SI)
- Liu, T., Merloni, A., Comparat, J., et al. 2022c, *A&A*, 661, A27 (eROSITA EDR SI)
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2010, *ApJS*, 187, 560
- Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, *Res. Astron. Astrophys.*, 15, 1095
- Maccacaro, 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
- McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, *The Messenger*, 154, 35
- Meisner, A. M., Lang, D., Schlafly, E. F., & Schlegel, D. J. 2019, *PASP*, 131, 124504
- Menzel, M.-L., et al. 2016, *MNRAS*, 457, 110
- Merloni, A. 2016, *Observing Supermassive Black Holes Across Cosmic Time: From Phenomenology to Physics*, eds. F. Haardt, V. Gorini, U. Moschella, A. Treves, & M. Colpi (Berlin: Springer), 905, 101
- Merloni, A., Predehl, P., Becker, W., et al. 2012, ArXiv e-prints [arXiv:1209.3114]
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, *PASJ*, 70, S1
- Naylor, T., Broos, P. S., & Feigelson, E. D. 2013, *ApJS*, 209, 30
- Noll, S., Mehlert, D., Appenzeller, I., et al. 2004, *A&A*, 418, 885
- Palanque-Delabrouille, N., Yèche, C., Myers, A. D., et al. 2011, *A&A*, 530, A122
- Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, *ApJS*, 234, 34
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, 663, 81
- Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, *A&A*, 647, A1
- Prevot, M. L., Lequeux, J., Maurice, E., Prevot, L., & Rocca-Volmerange, B. 1984, *A&A*, 132, 389
- Ruiz, A., Corral, A., Mountrichas, G., & Georgantopoulos, I. 2018, *A&A*, 618, A52
- Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
- Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, *ApJ*, 742, 61
- Salvato, M., Buchner, J., Budavári, T., et al. 2018, *MNRAS*, 473, 4937
- Salvato, M., Ilbert, O., & Hoyle, B. 2019, *Nat. Astron.*, 3, 212
- Schneider, P. C., Freund, S., Czesla, S., et al. 2022, *A&A*, 661, A6 (eROSITA EDR SI)
- Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, *ApJS*, 254, 24
- Shu, Y., Kopysov, S. E., Evans, N. W., et al. 2019, *MNRAS*, 489, 4741
- Simm, T., Saglia, R., Salvato, M., et al. 2015, *A&A*, 584, A106
- Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, *AJ*, 146, 32
- Sunyaev, R., Arefiev, V., Babushkin, V., et al. 2021, *A&A*, 656, A132
- Sutherland, W., & Saunders, W. 1992, *MNRAS*, 259, 413
- Taylor, M. B. 2005, *ASP Conf. Ser.*, 347, 29
- Taylor, M. B. 2006, *ASP Conf. Ser.*, 351, 666
- Toba, Y., Liu, T., Urrutia, T., et al. 2022, *A&A*, 661, A15 (eROSITA EDR SI)
- Vulic, N., Hornschemeier, A. E., Haberl, F., et al. 2022, *A&A*, 661, A16 (eROSITA EDR SI)
- Wenger, M., Ochsenein, F., Egret, D., et al. 2000, *A&AS*, 143, 9
- White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, *ApJ*, 475, 479
- Willis, J. P., Oguri, M., Ramos-Ceja, M. E., et al. 2021, *MNRAS*, 503, 5624
- Wolf, J., Nandra, K., Salvato, M., et al. 2021, *A&A*, 647, A5

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

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *AJ*, **120**, 1579

- 
- <sup>1</sup> Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr. 1, 85748 Garching, Germany  
e-mail: mara@mpe.mpg.de
- <sup>2</sup> Exzellenzcluster ORIGINS, Boltzmannstr. 2, 85748 Garching, Germany
- <sup>3</sup> Institute for Astronomy and Astrophysics, National Observatory of Athens, V. Paulou and I. Metaxa 11532, Greece
- <sup>4</sup> Dipartimento di Fisica e Astronomia "Augusto Righi", Università di Bologna, via Gobetti 93/2, 40129 Bologna, Italy
- <sup>5</sup> INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, 40129 Bologna, Italy
- <sup>6</sup> Department of Astronomy, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
- <sup>7</sup> Academia Sinica Institute of Astronomy and Astrophysics, Taipei 10617, Taiwan
- <sup>8</sup> Research Center for Space and Cosmic Evolution, Ehime University, 2-5 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan
- <sup>9</sup> Leibniz-Institut für Astrophysik Potsdam (AIP). An der Sternwarte 16. 14482 Potsdam, Germany
- <sup>10</sup> Universität Hamburg, Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany
- <sup>11</sup> Institute for Advanced Research, Nagoya University Furocho, Chikusa-ku, Nagoya, 464-8602 Japan
- <sup>12</sup> Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr 1, 81679 Munich, Germany
- <sup>13</sup> Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA
- <sup>14</sup> Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
- <sup>15</sup> Department of Physics, University of Connecticut, 2152 Hillside Road, Unit 3046, Storrs, CT 06269, USA
- <sup>16</sup> Department of Astronomy and Astrophysics, 525 Davey Lab, The Pennsylvania State University, University Park, PA 16802, USA
- <sup>17</sup> Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
- <sup>18</sup> Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA
- <sup>19</sup> Department of Physics and Yale Center for Astronomy and Astrophysics, Yale University, PO Box 208120, New Haven, CT 06520-8120, USA
- <sup>20</sup> Instituto de Astronomía Ensenada, Universidad Nacional Autónoma de México, AP 106, Ensenada 22800, Mexico
- <sup>21</sup> Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8578, Japan
- <sup>22</sup> Astronomical Institute, Tohoku University, Aramaki, Aoba-ku, Sendai, Miyagi 980-8578, Japan
- <sup>23</sup> Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA
- <sup>24</sup> Centro de Investigación en Astronomía, Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile
- <sup>25</sup> Apache Point Observatory and New Mexico State University, PO Box 59, Sunspot, NM 88349, USA
- <sup>26</sup> Centro de Astronomía (CITEVA), Universidad de Antofagasta, Avenida Angamos 601, Antofagasta 1270300, Chile

## Appendix A: Construction of the training, validation, and field samples

In the following, we describe the construction of the training, validation, and associated field samples we used to determine the different priors adopted by NWAY and ASTROMATCH to determine the counterparts and to assess the reliability of the association, presented in Section 5.

### Appendix A.1: Reference sample selected from the 3XMM-DR8 serendipitous source catalogue (the training sample)

We started with the 3XMM-DR8<sup>14</sup> catalogue of X-ray detections and estimated the 0.5-2 keV X-ray flux (and uncertainty) of each detection from the 0.5-1 and 1-2 keV band fluxes (and their uncertainties). We then selected only those detections that met all of the following X-ray quality criteria:

- i. have X-ray flux in the range probed by eFEDS ( $F_{0.5-2\text{keV}} > 2 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ , see figure 9 of Brunner et al. 2022),
- ii. have X-ray positions that are aligned with the optical frame and that have an uncertainty smaller than 1.5 arcsec,
- iii. have a signal-to-noise ratio for  $F_{0.5-2\text{keV}}$  that is greater than 10,
- iv. are consistent with being point-like at the resolution of *XMM-Newton*,
- v. have no close X-ray neighbours within 10 arcsec,
- vi. were not detected at the extreme off-axis angles,
- vii. were detected in *XMM-Newton* exposures of at least 5 ks,
- viii. were not labelled by the 3XMM pipeline as being confused, affected by high X-ray background, or flagged as being problematic for any reason.

We then excluded any X-ray detections that lie in parts of the sky that are not representative of a well-chosen extragalactic survey field such as eFEDS, or where the optical imaging catalogue (LS8) is likely to be saturated/unreliable (due to very bright stars). Specifically, we *exclude* any X-ray detections that:

- i. lie near the Galactic plane ( $|b| < 15 \text{ deg}$ ),
- ii. lie near the Large or Small Magellanic clouds or M31 (within 5, 3, and 1 degree radii, respectively),
- iii. lie within the disks of bright ( $B_T < 12$ ) well-resolved galaxies from de Vaucouleurs et al. (1991),
- vi. lie closer than 3 arcmin from any very bright star from the Yale Bright Star catalogue (Hoffleit 1964), or
- v. lie closer than 3 arcmin from any *Tycho-2* (Høg et al. 2000) star having  $B_T < 9$  or  $V_T < 9$ .

After applying these criteria, we were left with a sample of 36276 unique point-like X-ray sources with a median positional uncertainty 0.57 arcsec. The X-ray flux distribution of the 3XMM-DR8-based training sample is broadly similar to that of the eFEDS main sample. We find that 92% of the training sample have 0.5–2 keV fluxes in the range  $5 \times 10^{-15} - 1 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2}$ , and the median flux is  $1.7 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ . The equivalent metrics for eFEDS main sample are 91% and  $1.0 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ , respectively.

We then carried out a positional match of this X-ray training sample to the LS8, initially considering optical/IR objects that lay within 5 arcsec of the X-ray positions. We used NWAY

(Salvato et al. 2019) to carry out this cross-match, using only astrometric information and number densities (NWAY basic mode, i.e. without any magnitude or colour priors). We retained only the X-ray sources with very secure unique optical counterparts. Specifically, we required that we consider only X-ray sources with >90% probability of having an optical/IR counterpart, and only cases where the best optical/IR counterpart is at least nine times more probable than the next best possibility ( $p_{\text{any}} > 0.9$ ,  $p_{\text{i}} > 0.9$ )<sup>15</sup>. As before, we could afford to be very strict with these criteria, since we primarily cared about purity and not completeness. These cuts resulted in a 3XMM/LS8 reference sample of 20705 high-quality X-ray/OIR matches.

We selected a corresponding sample of non-X-ray emitting field objects from the LS8 using annular regions (15, 30 arcsec radii) around each of the 20705 reference sample positions. The field sample was further filtered to remove any object that lay within 15 arcsec of *any* 3XMM-DR8 source. This field sample contains just under 396 000 entries.

### Appendix A.2: Reference sample selected from the Chandra Source Catalogue v2.0 (the validation sample)

A supplementary X-ray/OIR reference sample was derived from the Chandra Source Catalogue v2.0<sup>16</sup>. We used the Web API to retrieve all CSC2 sources that satisfied the following X-ray quality criteria: i) Have  $F_{0.5-2\text{keV}} > 2 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$  (estimated from the standard CSC2 ‘s’ and ‘m’ bands), ii) have high significance  $> 6$ , iii) have a signal-to-noise ratio on  $F_{0.5-2\text{keV}}$  that is greater than 5, iv) have X-ray positions with 95% uncertainty ellipse radius smaller than 1.0 arcsec, v) are consistent with being point-like at the resolution of Chandra, vi) were detected in Chandra exposures of at least 1 ks, vii) were not labelled by the CSC pipeline as being confused, affected by readout streaks, or piled up. Exactly the same sky region filtering criteria were applied to the CSC-based sample as were used to filter the 3XMM-based reference sample (see Section A.1). These criteria resulted in a sample of 6066 X-ray sources.

We followed a similar process as before (Section A.1) to match the CSC2 sources to the LS8 catalogue, retaining only very secure matches (having  $p_{\text{any}} > 0.9$ ,  $p_{\text{i}} > 0.9$ ). This resulted in a CSC-based X-ray/OIR reference sample that contains 3415 objects.

## Appendix B: Templates used for photo-z

As discussed in the main text, a key ingredient for computing the photometric redshifts is the use of the most appropriate templates that can represent the population that is analysed. Here we list the templates we used for the EXT and PLIKE samples, respectively. For each of the templates, we provide the name of the model, the corresponding number in the catalog, and the reference to the paper that describes them.

### Appendix B.1: Templates used for PLIKE

The library of templates used for the PLIKE sample includes a mixture of SEDs from single objects and hybrids (different

<sup>14</sup> [http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR8/3XMM\\_DR8.html](http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR8/3XMM_DR8.html)

<sup>15</sup> in NWAY,  $p_{\text{any}}$  is the probability for each source in the primary catalogue (eFEDS in this case) to have a counterpart in the secondary catalogs; then, for each source in the secondary catalogues,  $p_{\text{i}}$  gives the probability to be the correct counterpart to the source in the primary catalogue (see more in the NWAY manual and Salvato et al. (2019)

<sup>16</sup> <https://cxc.cfa.harvard.edu/csc2/index.html>

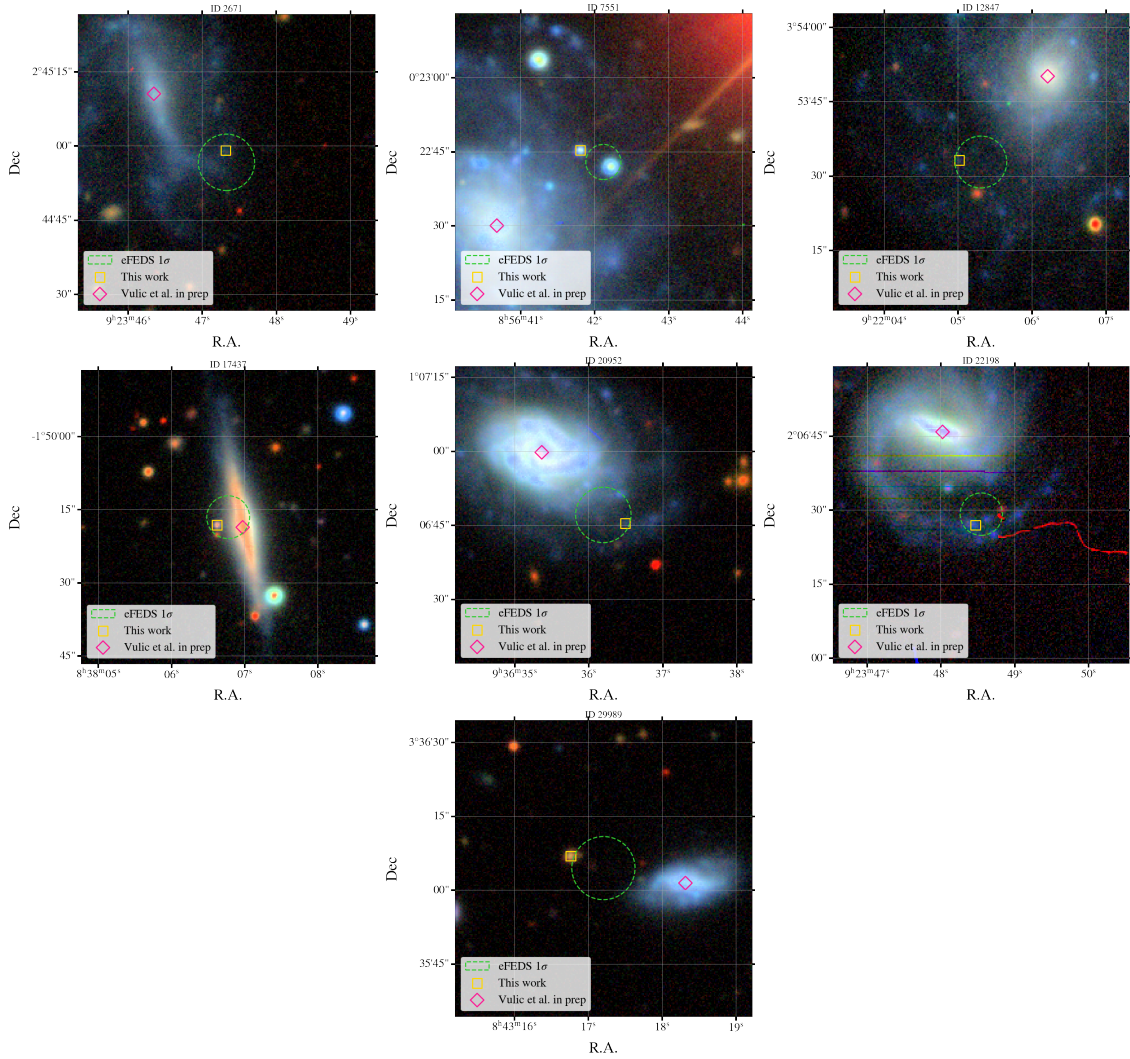


Fig. A.1: Seven sources for which the counterparts fall within an HECATE galaxy (see Vulic et al. 2022), but that are probably pointing to a background extragalactic source. The RGB images are  $1' \times 1'$ . In the cutouts we indicate the position of our proposed counterpart, the centre of the galaxy, the X-ray position, and the positional error.

relative contribution of host and nuclear component, as introduced in Salvato et al. (2009). Templates 1 and 12-17 and 20 are originally from Polletta et al. (2007). However, as templates 18 and 19 (originally from SDSS-V<sup>17</sup>, template 20 has been extended in the UV and presented in Salvato et al. (2009). The same template was then used to create hybrid templates by mixing it with normal galaxy templates (Noll et al. 2004) with a different degree of star formation, as presented by Ananna et al. (2017)<sup>18</sup>. With a similar procedure, hybrid templates 2-10 were constructed by combining an S0 and a QSO2 template, both from Polletta et al. (2007). The hybrids were originally presented in Salvato et al. (2009) and were successfully used in Salvato et al. (2009, 2011) and Marchesi et al. (2016) among others. Finally, templates 24-29 are from the recent work of Brown et al. (2019).

1. S0
2. S0\_10\_QSO2\_90
3. S0\_20\_QSO2\_80

<sup>17</sup> <http://classic.sdss.org/dr5/algorithms/spectemplates>

<sup>18</sup> The templates are slightly different than in Ananna et al. in the UV part.

4. S0\_30\_QSO2\_70
5. S0\_40\_QSO2\_60
6. S0\_50\_QSO2\_50
7. S0\_60\_QSO2\_40
8. S0\_70\_QSO2\_30
9. S0\_80\_QSO2\_20
10. S0\_90\_QSO2\_10
11. CB1\_0\_LOIII4
12. Sb
13. Spi4
14. M82
15. I22491
16. Sey18
17. Sey2
18. pl\_QSOH
19. pl\_QSO\_DR2\_029\_t0
20. pl\_TQSO1
21. s250\_10\_pl\_TQSO1\_90
22. s180\_30\_pl\_TQSO1\_70
23. s800\_20\_pl\_TQSO1\_80
24. 3C120
25. MRK110
26. NGC5548\_64.00\_NGC4138

27. PG0052p251
28. NGC4151
29. NGC3783\_00.50\_NGC4725

#### Appendix B.2: Templates used for EXT

The templates used to compute the photo-z for the EXT sample were almost entirely taken from Brown et al. (2019). The list of AGN templates includes the SED of single objects (the name of the template is self-explanatory) and hybrids constructed by combining at a different ratio the SED of two different galaxies and AGN. Additionally, we added two templates of elliptical galaxies from Polletta et al. (2007) and two newly created templates of type 1 AGN. We used the type 1 archetype presented in Comparat et al. (2020) that were obtained by stacking all type 1 sources that are counterparts to ROSAT/2RXS (Boller et al. 2016; Salvato et al. 2019) that had an SDSS spectra. The SED was then extended in the mid-infrared using the BQSO template presented in Polletta et al. (2007). BQSO is similar to TQSO, but with less enhanced mid-infrared flux. This new hybrid was then extended in the UV with various slopes.

1. 2MASXJ13000533+1632151
2. Ark564
3. F16156+0146
4. F2M1113+1244
5. Fairall9
6. H1821+643
7. IRAS\_11119+3257
8. Mrk110
9. Mrk1502
10. Mrk231
11. Mrk290
12. Mrk493
13. Mrk590
14. Mrk817
15. NGC3227\_Central\_00.50\_NGC4569
16. NGC3227\_Central\_01.00\_NGC4569
17. NGC3227\_Central\_02.00\_NGC4569
18. NGC3227\_Central\_04.00\_NGC4569
19. NGC3227\_Central\_08.00\_NGC4569
20. NGC3227\_Central\_16.00\_NGC4569
21. NGC3227\_Central\_32.00\_NGC4569
22. NGC3227\_Central\_64.00\_NGC4569
23. NGC3516\_Central\_00.50\_NGC4826
24. NGC3516\_Central\_01.00\_NGC4826
25. NGC3516\_Central\_02.00\_NGC4826
26. NGC3516\_Central\_04.00\_NGC4826
27. NGC3516\_Central\_08.00\_NGC4826
28. NGC3516\_Central\_16.00\_NGC4826
29. NGC3516\_Central\_32.00\_NGC4826
30. NGC3516\_Central\_64.00\_NGC4826
31. NGC3516\_Central
32. NGC3783\_Central\_01.00\_NGC4725
33. NGC3783\_Central\_02.00\_NGC4725
34. NGC3783\_Central\_04.00\_NGC4725
35. NGC3783\_Central\_08.00\_NGC4725
36. NGC3783\_Central\_16.00\_NGC4725
37. NGC3783\_Central\_32.00\_NGC4725
38. NGC3783\_Central\_64.00\_NGC4725
39. NGC4051\_Central\_00.50\_NGC3310
40. NGC4051\_Central\_00.50\_NGC4125
41. NGC4051\_Central\_00.50\_NGC4559
42. NGC4051\_Central\_01.00\_NGC3310
43. NGC4051\_Central\_01.00\_NGC4125
44. NGC4051\_Central\_01.00\_NGC4559
45. NGC4051\_Central\_02.00\_NGC3310
46. NGC4051\_Central\_02.00\_NGC4125
47. NGC4051\_Central\_02.00\_NGC4559
48. NGC4051\_Central\_04.00\_NGC3310
49. NGC4051\_Central\_04.00\_NGC4125
50. NGC4051\_Central\_04.00\_NGC4559
51. NGC4051\_Central\_08.00\_NGC3310
52. NGC4051\_Central\_08.00\_NGC4125
53. NGC4051\_Central\_08.00\_NGC4559
54. NGC4051\_Central\_16.00\_NGC3310
55. NGC4051\_Central\_16.00\_NGC4125
56. NGC4051\_Central\_16.00\_NGC4559
57. NGC4051\_Central\_32.00\_NGC3310
58. NGC4051\_Central\_32.00\_NGC4125
59. NGC4051\_Central\_32.00\_NGC4559
60. NGC4051\_Central\_64.00\_NGC3310
61. NGC4051\_Central\_64.00\_NGC4125
62. NGC4051\_Central\_64.00\_NGC4559
63. NGC4051\_Central
64. NGC4151\_Central\_00.50\_NGC4125
65. NGC4151\_Central\_00.50\_NGC4579
66. NGC4151\_Central\_01.00\_NGC3310
67. NGC4151\_Central\_01.00\_NGC4125
68. NGC4151\_Central\_01.00\_NGC4579
69. NGC4151\_Central\_02.00\_NGC3310
70. NGC4151\_Central\_02.00\_NGC4125
71. NGC4151\_Central\_02.00\_NGC4579
72. NGC4151\_Central\_04.00\_NGC3310
73. NGC4151\_Central\_04.00\_NGC4125
74. NGC4151\_Central\_04.00\_NGC4579
75. NGC4151\_Central\_08.00\_NGC3310
76. NGC4151\_Central\_08.00\_NGC4125
77. NGC4151\_Central\_08.00\_NGC4579
78. NGC4151\_Central\_16.00\_NGC3310
79. NGC4151\_Central\_16.00\_NGC4125
80. NGC4151\_Central\_16.00\_NGC4579
81. NGC4151\_Central\_32.00\_NGC3310
82. NGC4151\_Central\_32.00\_NGC4125
83. NGC4151\_Central\_32.00\_NGC4579
84. NGC4151\_Central\_64.00\_NGC3310
85. NGC4151\_Central\_64.00\_NGC4125
86. NGC4151\_Central\_64.00\_NGC4579
87. NGC5548\_Central\_00.50\_NGC4138
88. NGC5548\_Central\_01.00\_NGC4138
89. NGC5548\_Central\_02.00\_NGC4138
90. NGC5548\_Central\_04.00\_NGC4138
91. NGC5548\_Central\_08.00\_NGC4138
92. NGC5548\_Central\_16.00\_NGC4138
93. NGC5548\_Central\_32.00\_NGC4138
94. NGC5548\_Central\_64.00\_NGC4138
95. NGC5548\_Central
96. NGC5728
97. NGC7469
98. OQ\_530
99. PG0026+129
100. PG1415+451
101. PKS1345+12
102. Ton951
103. E114\_A\_0
104. E115\_A\_0
105. pl\_BQSO\_Co19\_sl-8
106. pl\_BQSO\_Co19\_sl-20

## Appendix C: Release of PDZ and SED fitting

For each primary (and secondary, in case it exists) counterpart to the eFEDS point sources, we make the redshift probability distribution and the SED fitting as in Figure C.1 available upon request.

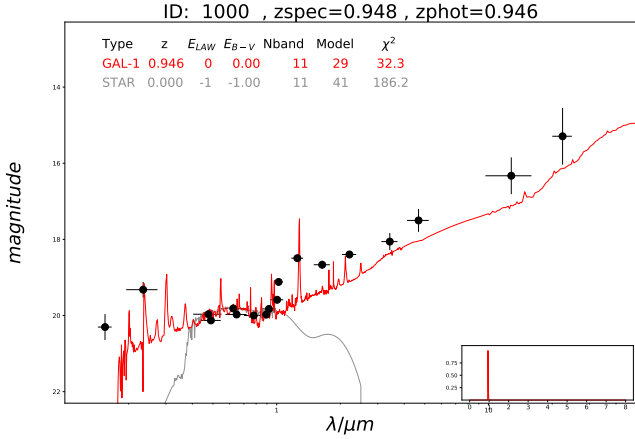


Fig. C.1: Example of SED fitting and redshift probability distribution (in the inset). The photometric points and their errors are indicated with black points. The best extragalactic and Galactic templates are shown with a solid line. The plot and the data to create the plots are available upon request.

## Appendix D: Column description of released catalogs

Most columns are common to both the main and hard sample. We list column descriptions for the main sample. In the few cases where the column descriptions differ, we report the corresponding description for the hard sample in brackets. The last column of the catalogue (ID\_MAIN) is present only in the hard catalogue.

- ERO\_Name:** eROSITA official source name (see Brunner et al.)
- ERO\_ID:** ID of eROSITA source in the main sample (from Brunner et al.) [ID of eROSITA source in the hard sample (from Brunner et al.)]
- ERO\_RA\_CORR;** J2000 Right Ascension of the eROSITA source (corrected) in degrees in the main sample (from Brunner et al.) [J2000 Right Ascension of the eROSITA source (corrected) in degrees in the hard sample (from Brunner et al.)]
- ERO\_Dec\_CORR;** J2000 Declination of the eROSITA source (corrected) in degrees in the main sample (from Brunner et al.) [J2000 Declination of the eROSITA source (corrected) in degrees in the hard sample (from Brunner et al.)]
- ERO\_RADEC\_ERR\_CORR:** eROSITA positional uncertainty (corrected) in arcsec from the main sample (from Brunner et al.) [eROSITA positional uncertainty (corrected) in arcsec from the hard sample (from Brunner et al.)]
- ERO\_ML\_FLUX:** 0.2-2.3 keV source flux in  $\text{erg cm}^{-2} \text{s}^{-1}$ , converted from the count rate assuming  $\text{ECF}=1.074\text{e}+12$  ( $\text{Gamma}=2.0$ ). [In the hard catalog, the column is called **ERO\_ML\_FLUX\_3**: 2.3-5 keV source flux in  $\text{erg cm}^{-2} \text{s}^{-1}$ , converted from the count rate assuming  $\text{ECF}=1.147\text{e}+11$  ( $\text{Gamma}=2.0$ ).] See Brunner et al.
- ERO\_ML\_FLUX\_ERR:** 0.2-2.3 keV source flux error ( $1\sigma$ ) in  $\text{erg cm}^{-2} \text{s}^{-1}$ . [In the hard catalog, the column is called **ERO\_ML\_FLUX\_ERR\_3**: 2.3-5 keV error ( $1\sigma$ ) in  $\text{erg cm}^{-2} \text{s}^{-1}$ ]. See Brunner et al.
- ERO\_DET\_LIKE:** 0.2-2.3 keV detection likelihood measured by PSF-fitting. [In the hard catalog, the column is called **ERO\_DET\_LIKE\_3**: 2.3-5 keV detection likelihood measured by PSF fitting]. See Brunner et al.
- ERO\_inArea90:** true if in the 0.2-2.3keV  $\text{exp}>500\text{s}$  region, which comprises 90% of the area (from Brunner et al.).
- CTP\_LS8\_UNIQUE\_OBJID:** LS8 unique identifier for the proposed counterpart to the eROSITA source (Expression:  $\text{toString}(\text{LS8\_BRICKID})+"_" + \text{toString}(\text{LS8\_OBJID})$ ).
- CTP\_LS8\_RA:** J2000 Right Ascension of the LS8 counterpart in degrees.
- CTP\_LS8\_Dec:** J2000 Declination of the LS8 counterpart in degrees.
- Dist\_CTP\_LS8\_ERO:** Separation between selected counterpart and eROSITA (corrected) position in arcsec.
- CTP\_NWAY\_LS8\_UNIQUE\_OBJID:** LS8 unique identifier of the best LS8 counterpart from NWAY (Expression:  $\text{toString}(\text{LS8\_BRICKID})+"_" + \text{toString}(\text{LS8\_OBJID})$ ).
- CTP\_NWAY\_LS8\_RA:** J2000 Right Ascension of the LS8 counterpart from NWAY in degrees.
- CTP\_NWAY\_LS8\_Dec:** J2000 Declination of the LS8 counterpart from NWAY in degrees.
- CTP\_NWAY\_dist\_bayesfactor:** Logarithm of the ratio of the prior and posterior from separation, positional error, and number density (see the appendix in Salvato et al 2018 for clarification).
- CTP\_NWAY\_dist\_post:** Distance posterior probability comparing this association vs. no association (see the appendix in Salvato et al 2018 for clarification).
- CTP\_NWAY\_p\_single:** Same as **dist\_post**, but weighted by the prior (see the appendix in Salvato et al 2018 for clarification).
- CTP\_NWAY\_p\_any:** For each entry in the X-ray catalogue, the probability that there is a counterpart (see the appendix in Salvato et al 2018 for clarification).
- CTP\_NWAY\_p\_i:** Relative probability of the eROSITA/LS8 match (see the appendix in Salvato et al 2018 for clarification).
- Dist\_NWAY\_LS8\_ERO:** Separation between the eROSITA position and the LS8 counterpart from NWAY in arcsec.
- CTP\_MLR\_LS8\_UNIQUE\_OBJID:** LS8 unique identifier of the LS8 counterpart from ASTROMATCH (Expression:  $\text{toString}(\text{LS8\_BRICKID})+"_" + \text{toString}(\text{LS8\_OBJID})$ ).
- CTP\_MLR\_LS8\_RA :** J2000 Right Ascension of LS8 counterpart from ASTROMATCH in degrees.
- CTP\_MLR\_LS8\_Dec:** J2000 Declination of LS8 counterpart from ASTROMATCH in degrees.
- CTP\_MLR\_LR\_BEST:** Likelihood Ratio value from ASTROMATCH.
- CTP\_MLR\_REL\_BEST:** Reliability of the identification from ASTROMATCH.
- Dist\_MLR\_LS8\_ERO:** Separation between the eROSITA position and the LS8 counterpart from ASTROMATCH in arcsec.
- CTP\_SAME:** Comparison NWAY/MLR: true if the counterpart selected by the two method is the same.

30. **CTP\_MLR**: Comparison NWay/MLR: true if the counterpart from NWay(MLR) has `p_any(LR_BEST)` below(above) threshold.
31. **CTP\_Hamstar**: Match to HamStar: 1=same counterpart, 0=different counterpart, -99=no HamStar (Schneider et al.).
32. **CTP\_Hamstar\_p\_stellar**: Probability of association from HamStar.
33. **Dist\_CTP\_Hamstar**: Separation between the counterpart proposed by HamStar and the counterpart selected in this work.
34. **CTP\_quality**: counterpart quality: 4=best, 3=good, 2=with secondary, 1/0=unreliable (see paper)
35. **GaiaEDR3\_ID**: ID in Gaia EDR3 source catalog.
36. **GaiaEDR3\_parallax**: Parallax from Gaia EDR3 in mas.
37. **GaiaEDR3\_parallax\_error**: Parallax error from Gaia EDR3 in mas.
38. **GaiaEDR3\_parallax\_over\_error**: Parallax/Parallax error. (a ratio >5 define a SECURE GALACTIC counterpart).
39. **GaiaEDR3\_pmra**: Proper motion in RA from Gaia EDR3.
40. **GaiaEDR3\_pmra\_error**: Error on Proper motion in RA from Gaia EDR3.
41. **GaiaEDR3\_pmdec**: Proper motion in Dec from Gaia EDR3.
42. **GaiaEDR3\_pmdec\_error**: Error on Proper motion in Dec from Gaia EDR3.
43. **GaiaEDR3\_phot\_g\_mean\_mag**: g band magnitude (VEGA) from Gaia EDR3.
44. **GaiaEDR3\_phot\_g\_mean\_mag\_error**: Error g band magnitude (VEGA) from Gaia EDR3.
45. **GaiaEDR3\_phot\_bp\_mean\_mag**: bp band magnitude (VEGA) from Gaia EDR3.
46. **GaiaEDR3\_phot\_bp\_mean\_mag\_error**: Error bp band magnitude (VEGA) from Gaia EDR3.
47. **GaiaEDR3\_phot\_rp\_mean\_mag**: rp band magnitude (VEGA) from Gaia EDR3.
48. **GaiaEDR3\_phot\_rp\_mean\_mag\_error**: Error rp band magnitude (VEGA) from Gaia EDR3.
49. **FUV**: Galex Far UV magnitude (AB magnitude).
50. **FUV\_ERR**: Galex Far UV magnitude error (AB magnitude).
51. **NUV**: Galex Near UV magnitude (AB magnitude).
52. **NUV\_ERR**: Galex Near UV magnitude error (AB magnitude).
53. **KiDS\_u**: KiDS u-band magnitude (AB magnitude).
54. **KiDS\_u\_ERR**: KiDS u-band magnitude error (AB magnitude).
55. **KiDS\_g**: KiDS g-band magnitude (AB magnitude).
56. **KiDS\_g\_ERR**: KiDS g-band magnitude error (AB magnitude).
57. **KiDS\_r**: KiDS r-band magnitude (AB magnitude).
58. **KiDS\_r\_ERR**: KiDS r-band magnitude error (AB magnitude).
59. **KiDS\_i**: KiDS i-band magnitude (AB magnitude).
60. **KiDS\_i\_ERR**: KiDS i-band magnitude error (AB magnitude).
61. **omegac\_z**: OmegaCAM z-band magnitude (AB magnitude).
62. **omegac\_z\_ERR**: OmegaCAM z-band magnitude error (AB magnitude).
63. **HSC\_g**: HSC g-band magnitude (AB magnitude).
64. **HSC\_g\_ERR**: HSC g-band magnitude error (AB magnitude).
65. **HSC\_r**: HSC r-band magnitude (AB magnitude).
66. **HSC\_r\_ERR**: HSC r-band magnitude error (AB magnitude).
67. **HSC\_r2**: HSC r2-band magnitude (AB magnitude).
68. **HSC\_r2\_ERR**: HSC r2-band magnitude error (AB magnitude).
69. **HSC\_i**: HSC i-band magnitude (AB magnitude).
70. **HSC\_i\_ERR**: HSC i-band magnitude error (AB magnitude).
71. **HSC\_i2**: HSC i2-band magnitude (AB magnitude).
72. **HSC\_i2\_ERR**: HSC i2-band magnitude error (AB magnitude).
73. **HSC\_z**: HSC z-band magnitude (AB magnitude).
74. **HSC\_z\_ERR**: HSC z-band magnitude error (AB magnitude).
75. **HSC\_Y**: HSC Y-band magnitude (AB magnitude).
76. **HSC\_Y\_ERR**: HSC Y-band magnitude error (AB magnitude).
77. **VIKING\_z**: VISTA/VIKING z-band magnitude (AB magnitude).
78. **VIKING\_z\_ERR**: VISTA/VIKING z-band magnitude error (AB magnitude).
79. **VIKING\_Y**: VISTA/VIKING Y-band magnitude (AB magnitude).
80. **VIKING\_Y\_ERR**: VISTA/VIKING Y-band magnitude error (AB magnitude).
81. **VIKING\_J**: VISTA/VIKING J-band magnitude (AB magnitude).
82. **VIKING\_J\_ERR**: VISTA/VIKING J-band magnitude error (AB magnitude).
83. **VIKING\_H**: VISTA/VIKING H-band magnitude (AB magnitude).
84. **VIKING\_H\_ERR**: VISTA/VIKING H-band magnitude error (AB magnitude).
85. **VIKING\_Ks**: VISTA/VIKING Ks-band magnitude (AB magnitude).
86. **VIKING\_Ks\_ERR**: VISTA/VIKING Ks-band magnitude error (AB magnitude).
87. **W1**: LS8/Wise W1 magnitude (AB magnitude).
88. **W1\_ERR**: LS8/Wise W1 magnitude error (AB magnitude).
89. **W2**: LS8/Wise W2 magnitude (AB magnitude).
90. **W2\_ERR**: LS8/Wise W2 magnitude error (AB magnitude).
91. **W3**: LS8/Wise W3 magnitude (AB magnitude).
92. **W3\_ERR**: LS8/Wise W3 magnitude error (AB magnitude).
93. **W4**: LS8/Wise W4 magnitude (AB magnitude).
94. **W4\_ERR**: LS8/Wise W4 magnitude error (AB magnitude).
95. **LS8\_g**: LS8 g-band magnitude (AB magnitude).
96. **LS8\_g\_ERR**: LS8 g-band magnitude error (AB magnitude).
97. **LS8\_r**: LS8 r-band magnitude (AB magnitude).
98. **LS8\_r\_ERR**: LS8 r-band magnitude error (AB magnitude).
99. **LS8\_z**: LS8 z-band magnitude (AB magnitude).
100. **LS8\_z\_ERR**: LS8 z-band magnitude error (AB magnitude).
101. **VHS\_Y**: VISTA/VHS Y-band magnitude (AB magnitude).
102. **VHS\_Y\_ERR**: VISTA/VHS Y-band magnitude error (AB magnitude).
103. **VHS\_H**: VISTA/VHS H-band magnitude (AB magnitude).
104. **VHS\_H\_ERR**: VISTA/VHS H-band magnitude error (AB magnitude).
105. **VHS\_Ks**: VISTA/VHS Ks-band magnitude (AB magnitude).
106. **VHS\_Ks\_ERR**: VISTA/VHS Ks-band magnitude error (AB magnitude).
107. **HCS\_g\_diff**: Difference between PSF and Kron magnitude in HSC g band (AB magnitude).
108. **HCS\_r\_diff**: Difference between PSF and Kron magnitude in HSC r band (AB magnitude).

109. **HCS\_i\_diff**: Difference between PSF and Kron magnitude in HSC i band (AB magnitude).
110. **HCS\_z\_diff**: Difference between PSF and Kron magnitude in HSC z band (AB magnitude).
111. **HCS\_opt\_extended**: Extension in HSC griz bands. 1=extended; -99=data missing 0=other from Aihara et al. (2018).
112. **CTP\_LS8\_phot\_flag**: Flag for LS8 photometry: true when the source has simultaneously g,r,z,w1 photometry in LS8.
113. **CTP\_LS8\_Type**: Morphological model from LS8.
114. **in\_KiDS\_flag**: Flag for KiDS coverage: 1: Source is in KiDS area; 0: otherwise.
115. **in\_HSC\_flag**: Flag for HSC coverage: 1: Source is in HSC area as from Aihara et al. (2018); 0: otherwise.
116. **SPECZ\_RA**: Right Ascension (degrees) of the spectroscopic redshift entry in the original catalogue from which it was taken.
117. **SPECZ\_Dec**: Declination (degrees) of the spectroscopic redshift entry in the original catalogue from which it was taken.
118. **SPECZ\_Redshift**: Spectroscopic redshift from the original catalog.
119. **SPECZ\_NORMQ**: Normalised quality of spectroscopic redshift: 3=secure, 2=not secure, 1=unreliable redshift/bad spectrum, and -1= Blazar candidate
120. **SPECZ\_Origin**: Catalogue that provided this spectroscopic redshift.
121. **SPECZ\_Original\_ID**: Identifier of this spectroscopic redshift entry in the original catalogue from which it was taken.
122. **SPEC\_Gal\_flag** : True when the CTP has a reliable redshift above 0.002 (boolean)
123. **SPEC\_Star\_flag** : True when the CTP has a reliable redshift below 0.002 (boolean)
124. **CTP\_Classification** : Counterpart classification: SECURE/LIKELY GALACTIC/EXTRAGALACTIC, as from the flowchart (see paper).
125. **PHZ\_LePHARE\_zphot** : Photo-z from LE PHARE but set to 0 for GALACTIC sources.
126. **PHZ\_LePHARE\_zl68** : LE PHARE zphot min at  $1\sigma$ .
127. **PHZ\_LePHARE\_zu68** : LE PHARE zphot max at  $1\sigma$ .
128. **PHZ\_LePHARE\_zl90** : LE PHARE zphot min at  $2\sigma$ .
129. **PHZ\_LePHARE\_zu90** : LE PHARE zphot max at  $2\sigma$ .
130. **PHZ\_LePHARE\_zl99** : LE PHARE zphot min at  $3\sigma$ .
131. **PHZ\_LePHARE\_zu99** : LE PHARE zphot max at  $3\sigma$ .
132. **PHZ\_LePHARE\_chi** : LE PHARE chi2 value for best-fitting galaxy/AGN template.
133. **PHZ\_LePHARE\_ModelAGN** : LE PHARE best template fitting the data.
134. **PHZ\_LePHARE\_extlaw** : LE PHARE extinction law applied to the template: Prevot (1), or none (0).
135. **PHZ\_LePHARE\_ebv** : LE PHARE applied to the template.
136. **PHZ\_LePHARE\_pdz** : LE PHARE probability distribution. The photo-z is more reliable when the value is high.
137. **PHZ\_LePHARE\_nband** : LE PHARE number of bands used to compute the photo-z.
138. **PHZ\_LePHARE\_zp2** : LE PHARE second-best photo-z from LE PHARE, if it exists.
139. **PHZ\_LePHARE\_chi2\_2** : LE PHARE chi2 value for second-best fitting template, if it exists.
140. **PHZ\_LePHARE\_ModelAGN\_2** : LE PHARE second-best template fitting the data, if it exists.
141. **PHZ\_LePHARE\_pdz2** : LE PHARE probability distribution for the secondary solution, if it exists.
142. **PHZ\_DNNz\_zphot**: Photo-z from DNNZ (from Nishizawa et al.), but set to 0 for GALACTIC sources.
143. **PHZ\_DNNz\_zl68** : DNNZ  $1\sigma$ . min error on photo-z.
144. **PHZ\_DNNz\_zu68** : DNNZ  $1\sigma$ . max error on photo-z.
145. **PHZ\_DNNz\_zl95** : DNNZ  $2\sigma$ . min error on photo-z.
146. **PHZ\_DNNz\_zu95** : DNNZ  $2\sigma$ . max error on photo-z.
147. **CTP\_REDSHIFT** : Final redshift: zspec (NORMQ=3) when available, otherwise, photo-z from LE PHARE.
148. **CTP\_REDSHIFT\_GRADE** : In a range from 5 (spectroscopy) to 0 (unreliable photo-z) (see text for details).
149. **CLUSTER\_CLASS**: In a range from 5 to 1: 5=most likely a cluster; 1= not a cluster (see text for details).
150. **CTP\_CLASS** : same as CTP\_Classification, but with numbers: 3: SECURE EXTRAGALACTIC, 2: LIKELY EXTRAGALACTIC, 1: SECURE GALACTIC, and 0: LIKELY GALACTIC.
151. **ID\_MAIN**: [column present only in the hard sample: the source ID in the main catalogue for the sources in common.]