

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society ©: 2020 The Author Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# The next-generation X-ray galaxy survey with *eROSITA*

Antara R. Basu-Zych<sup>1,2,3</sup>★, Ann E. Hornschemeier<sup>1,3</sup>, Frank Haberl<sup>4</sup>, Neven Vulic<sup>1,5</sup>, Jörn Wilms<sup>6</sup>,  
Andreas Zezas<sup>7,8,9</sup>, Konstantinos Kovalakas<sup>1,9</sup>, Andrew Ptak<sup>1,3</sup> and Thomas Dauser<sup>1,6</sup>

<sup>1</sup>NASA Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA

<sup>2</sup>Department of Physics, University of Maryland Baltimore County, Baltimore, MD 21250, USA

<sup>3</sup>Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

<sup>4</sup>Max-Planck-Institut für extraterrestrische Physik, Gießenbachstraße 1, D-85748 Garching, Germany

<sup>5</sup>Department of Astronomy and Center for Space Science and Technology (CRESST), University of Maryland, College Park, MD 20742-2421, USA

<sup>6</sup>Dr. Karl Remeis-Observatory and Erlangen Centre for Astroparticle Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Sternwartstr. 7, D-96049 Bamberg, Germany

<sup>7</sup>Physics Department and Institute of Theoretical and Computational Physics, University of Crete, GR-71003 Heraklion, Crete, Greece

<sup>8</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>9</sup>Institute of Astrophysics, Foundation for Research and Technology-Hellas, GR-71110 Heraklion, Greece

Accepted 2020 July 24. Received 2020 June 29; in original form 2020 March 2

## ABSTRACT

*eROSITA*, launched on 2019 July 13, will be completing the first all-sky survey in the soft and medium X-ray band in nearly three decades. This 4-yr survey, finishing in late 2023, will present a rich legacy for the entire astrophysics community and complement upcoming multiwavelength surveys (with, e.g. the Large Synoptic Survey Telescope and the Dark Energy Survey). Besides the major scientific aim to study active galactic nuclei (AGNs) and galaxy clusters, *eROSITA* will contribute significantly to X-ray studies of normal (i.e. not AGN) galaxies. Starting from multiwavelength catalogues, we measure star formation rates and stellar masses for 60 212 galaxies constrained to distances of 50–200 Mpc. We chose this distance range to focus on the relatively unexplored volume outside the local Universe, where galaxies will be largely spatially unresolved and probe a range of X-ray luminosities that overlap with the low luminosity and/or highly obscured AGN population. We use the most recent X-ray scaling relations as well as the on-orbit *eROSITA* instrument performance to predict the X-ray emission from XRBs and diffuse hot gas and to perform both an analytic prediction and an end-to-end simulation using the mission simulation software, SIXTE. We consider potential contributions from hidden AGN and comment on the impact of normal galaxies on the measurement of the faint end of the AGN luminosity function. We predict that the *eROSITA* 4-yr survey, will detect  $\gtrsim 15\,000$  galaxies ( $3\sigma$  significance) at 50–200 Mpc, which is  $\sim 100 \times$  more normal galaxies than detected in any X-ray survey to date.

**Key words:** surveys – galaxies: statistics – X-rays: galaxies.

## 1 INTRODUCTION

It has now been more than two decades since X-ray surveys introduced normal galaxies as a population of X-ray emitters (e.g. with the *Einstein* Observatory and the *ROSAT* all-sky survey, Fabbiano 1989; Boller et al. 1992; Fabbiano, Kim & Trinchieri 1992; Zimmermann et al. 2001; Tajer et al. 2004). The X-ray emission from normal galaxies arises chiefly from accretion on to compact objects in binary systems and from the hot phase of the interstellar medium (ISM). It is important to study normal galaxies at X-ray energies for a variety of reasons. The dominant accreting compact object population, in the absence of an accreting supermassive black hole (SMBH), are neutron star (NS) and stellar-mass black hole (BH) populations, which uniquely trace the endpoint of massive star formation. Observations of BH and NS in accreting X-ray binaries (XRBs) tell us how they form and evolve and additionally provide understanding of the binary phase of stellar evolution. Further, the collective X-ray output from XRBs can rival that of accreting

supermassive BH (SMBH or Active Galactic Nuclei, AGNs) at the critical epochs of reionization and Cosmic Dawn ( $6 \lesssim z \lesssim 20$ ) when the first galaxies in the Universe were forming (Fragos et al. 2013; Mesinger, Ewall-Wice & Hewitt 2014; Pacucci et al. 2014). These lower mass, accreting systems effectively ‘outshine’ their SMBH counterparts, playing a possibly significant role in heating the primordial intergalactic medium (IGM). However given the difficulty of directly observing X-ray emission at high-redshifts, it is local studies of NS and BH emission that gives us the best chance to characterize this emission.

The hot diffuse gas in galaxies is similarly important to understand. The hot gaseous haloes of early-type galaxies embeds information regarding the formation and evolution of the galaxy. The thermal structure of the ISM, in which X-ray studies play a key role, also gives key information about various physical processes affecting galaxy evolution (e.g. see Kim & Fabbiano 2015). Hot gas properties are expected to differ based on e.g. mergers, ram-pressure stripping and both AGN and stellar feedback. Although there have been very detailed studies of small samples of galaxies in the local Universe, it has been impossible to do a large statistical study of the hot gas properties of early-type galaxies.

★ E-mail: antara.r.basu-zych@nasa.gov

All of our current understanding, however, is based on very small numbers of galaxies as compared to the powerhouse wide-field surveys in the optical/IR such as the Sloan Digital Sky Survey (SDSS). We certainly know that by number, X-ray emitting normal galaxies vastly outnumber galaxies harbouring AGNs, particularly at faint X-ray fluxes. Comparing the number counts for AGN with those for galaxies (henceforth, we refer to normal galaxies as galaxies), from the 4 Ms *Chandra* Deep Field South, Lehmer et al. (2012) find that galaxies become the dominant population at the faint end ( $S[0.5\text{--}10\text{ keV}] < 10^{-16}\text{ erg s}^{-1}\text{ cm}^{-2}$ ; see also Bauer et al. 2004). We are on the threshold of a new generation of wide-field X-ray surveys, where galaxy samples detected in the X-ray band will finally reach much larger numbers (thousands of galaxies versus 10s to 100s).

The current state of the art for X-ray surveys of normal galaxies with *Chandra* and *XMM-Newton* may be broadly divided into pencil-beam (contiguous or near-contiguous fields) and serendipitous (non-contiguous fields) with the largest areas surveys being in the tens of square degrees.<sup>1</sup> Both the serendipitous (Georgakakis et al. 2004; Georgantopoulos, Georgakakis & Koulouridis 2005) and the pencil beam surveys (Hornschemeier et al. 2003; Bauer et al. 2004; Georgakakis et al. 2006; Kim et al. 2006) have allowed the comparison of the X-ray emission to that at other wavelengths, such as infrared and optical (i.e.  $f_X/f_{\text{opt}}$ ), to aid in classifying AGN versus normal galaxies, and revealed important scaling relations of X-ray luminosity with global galaxy properties in normal galaxies (i.e. star formation rate, SFR, and stellar mass,  $M_*$ ; David, Jones & Forman 1992; Ranalli, Comastri & Setti 2003; Norman et al. 2004). We have learned how accreting binary populations evolve over billions of years of cosmic history, demonstrating both progenitor paths for eventual gravitational wave mergers as may be detected by LIGO (Maccarone et al. 2014) as well as showing that stellar-origin compact objects likely have significant impact on the heating of primordial IGM at  $z > 10$  (Mesinger, Ferrara & Spiegel 2013; Mesinger et al. 2014; Pacucci et al. 2014; Das et al. 2017; Greig & Mesinger 2017).

In this paper, we used the updated X-ray galaxy scaling relations with SFR and stellar mass, combined with newly available galaxy catalogues, to simulate X-ray observations of galaxies in a ‘bottom-up’ approach to predict full-sky X-ray galaxy survey numbers. We predicted which galaxies could be detected by the *eROSITA* X-ray survey, which is operating on board the Spectrum-Roentgen-Gamma mission that launched on 2019 July 13 (Merloni et al. 2012; Predehl et al. 2016).

*eROSITA* offers the next major step forward for studying galaxies at X-ray energies. *eROSITA* will scan the entire sky eight times over four years, using an array of seven aligned mirror modules, operating in the 0.2–10 keV energy range and covering a  $\sim 1^\circ$  diameter field-of-view (FOV; Merloni et al. 2012). The on-axis point-spread function (PSF) at 1.5 keV is about 15 arcsec (half energy width, HEW) and degrades with off-axis angle, resulting in an average of 28 arcsec over the FOV. With a point source X-ray survey sensitivity (after four years) in the 0.5–2 keV band of  $\sim 1.1 \times 10^{-14}\text{ erg s}^{-1}\text{ cm}^{-2}$  for an average all-sky exposure<sup>2</sup> of  $\sim 2\text{ ks}$ , *eROSITA* will deliver the largest catalogue of galaxies detected in X-rays to date.

Approximately a decade ago, Prokopenko & Gilfanov (2009) made a prediction of the normal galaxy population that might be seen by *eROSITA* surveys. They estimated 15 000–20 000 ( $\sim 8\,400$  early-

type and 7 000–10 000 late-type) galaxies would be detected, using a ‘top-down’ technique starting from the best knowledge, at that time, of the galaxy X-ray luminosity function (XLF) and its evolution. Since then, there has been great improvement in our knowledge of X-ray scaling relationships, in the quality of galaxy catalogues (permitting a ‘bottom-up’ approach) and of course improvement in our knowledge of the plans for the *eROSITA* surveys.

Specifically, there have been major improvements in our understanding of the X-ray scaling relationships for galaxies based on properties such as their SFR/stellar mass (Mineo, Gilfanov & Sunyaev 2012a; Basu-Zych et al. 2013a; Lehmer et al. 2016). There has also been major maturation of galaxy catalogues in the local Universe, meaning that galaxy SFRs and stellar masses are well constrained for a significant population of relatively nearby ( $D \lesssim 200\text{ Mpc}$ ) galaxies. These two major advances mean that a bottom-up approach to modelling X-ray emission from galaxies is now possible. In our paper, we used such an approach to explore the capabilities of future X-ray survey experiments for doing normal galaxy science.

We focused our study on galaxies at distances 50–200 Mpc, where we expect *eROSITA* observations will provide significant gains over available X-ray surveys. This distance range represents the relatively unexplored volume for studying X-ray emission from normal galaxies outside the more local Universe. Our study included both early- and late-type galaxies, and used updated galaxy catalogues, revised scaling relations and state-of-the-art X-ray simulations with current *eROSITA* response files to provide the most rigorous estimates. We describe our sample selection throughout Section 2, including our methods for measuring SFRs and stellar masses in Section 2.5 and screening against AGN in Section 2.6. In Section 3, we outline our techniques for simulating *eROSITA* galaxies, using an analytic method (Section 3.1) and also an end-to-end simulation with the Simulation of X-ray Telescopes (SIXTE) software (Section 3.4). We present our results and conclusions in Sections 4 and 5, respectively. Throughout the paper, we assume the Kroupa (2001) initial mass function (IMF) when measuring SFRs and stellar masses, and adopt  $\Lambda\text{CDM}$  cosmology with the following parameters:  $H_0 = 70\text{ Mpc km s}^{-1}$ ,  $\Omega_M = 0.3$ ,  $\Omega_\Lambda = 0.7$ .

## 2 THE CATALOGUE

We drew our sample from the current version of the galaxy catalogue of Kovlakas et al. (in preparation), henceforth referred to as the Heraklion Extragalactic CATalogueE (HECATE), that contains all nearby galaxies ( $D \lesssim 200\text{ Mpc}$ ) from the HyperLEDA data base (Makarov et al. 2014), along with their redshifts and redshift-independent distances, where available, carefully accounting for various potential sources of errors (including infall to the Great Attractor and Shapley supercluster). The HECATE catalogue imposed a Virgo-infall corrected radial velocity cut of  $14\,000\text{ km s}^{-1}$  (corresponding to  $\approx 200\text{ Mpc}$ ) on sources marked with ‘G’,<sup>3</sup> to indicate a galaxy in the HyperLEDA catalogue. We note that the ‘G’ object class designation not only excludes stars, star clusters, etc., but also reduces contamination from ‘parts of galaxies’, and ‘multiple galaxies’, which may appear multiply listed in other catalogues (e.g. NGC 4038/NGC 4039 may appear with a variety of nomenclature: NGC 4038/NGC 4039 or NGC 4038 and NGC 4039). Therefore, restricting the sample to those with the ‘G’ criterion provides better cross-matching with other catalogues. The ‘G’ criterion also allowed us to avoid regions where the X-ray emission may be dominated by

<sup>1</sup> There has also been a nearly all-sky survey, the *XMM-Newton* Slew Survey, however this is at a fairly bright X-ray flux limit that is appropriate for studying relatively luminous and nearby AGN (e.g. Dwelly et al. 2017)

<sup>2</sup> Some areas will have deeper exposure, such as the ecliptic poles; see Fig. 13

<sup>3</sup> <http://leda.univ-lyon1.fr/leda/param/objtype.html>

**Table 1.** Tracking changes in the sample size with each step in the analysis. The columns provide (1) row number for easy reference, (2) the analysis step, (3) the section that details the method, (4) the sample size within the HECATE-SDSS and (5) HECATE-IRAS samples, and (6) the combined sample size at the end of the step. Note that numbers within parentheses signify the total number which includes duplicate sources between the HECATE-SDSS and HECATE-IRAS samples. However, after Section 2.3 we account for these within the HECATE-SDSS sample only and therefore the numbers from this point reflect the unique numbers of galaxies. The \* signifies those numbers which still count the duplicates.

Row (1)	Analysis step (2)	Reference section(s) (3)	HECATE-SDSS # (4)	HECATE-IRAS # (5)	(Total) Unique # (6)
(1)	Parent sample	2.1 & 2.2	67 875	17 719*	(85 594) 81 024
(2)	$50 \leq D \leq 200$ (all)	2.4	67 159	14 646*	(81 805)
(3)	$50 \leq D \leq 200$ (duplicates removed)	–	67 159	10 186	77 345
(4)	$50 \leq D \leq 200 \cap \text{SFR}$ (Eq 1)	2.5.1	47 831	5 535	61 067
(5)	$50 \leq D \leq 200 \cap \text{SFR}$ ('unWISE')	–	7 701	...	–
(6)	$50 \leq D \leq 200 \cap \text{SFR}$ (Total)	–	55 532	5 535	–
(7)	$50 \leq D \leq 200 \cap M_*$ (Eq 2)	2.5.2	46 608	3 398	76 777
(8)	$50 \leq D \leq 200 \cap M_*$ (SED-fit)	–	20 241	...	–
(9)	$50 \leq D \leq 200 \cap M_*$ (extrapolated $g-r$ )	–	...	6 530	–
(10)	$50 \leq D \leq 200 \cap M_*$ (Total)	–	66 849	9 928	–
(11)	$50 \leq D \leq 200 \cap M_* \cap \text{SFR}$	2.5.2	55 312	5 476	60 788
(12)	$50 \leq D \leq 200 \cap M_* \cap \text{SFR} \cap \text{AGN Screening}$	2.6	54 736	5 476	60 212

the hot gas related to e.g. the intracluster or intragroup environment rather than intragalaxy X-ray emission from XRBs and/or the ISM. Of course, X-ray emission from AGN is also a concern, and we discuss our selection against AGN in Section 2.6 and our treatment of potential contamination in more detail in Section 3.5.

In addition, the HECATE galaxies were required to have coordinates with astrometric precision of  $< 10$  arcsec (note that  $\sim 92$  per cent of the objects in the HyperLEDA catalogue have precisions of  $< 1$  arcsec) and were rejected if missing astrometric precisions. The method mentioned above yielded a sample of 163 648 HECATE galaxies.

As we describe in the following two subsections, we used two subsets of the parent HECATE catalogue: the HECATE-SDSS sample (67 875 galaxies), which includes galaxy properties from the *GALEX*–SDSS–WISE Legacy Catalog (GSWLC; Salim et al. 2016), and the HECATE-IRAS sample (17 719 galaxies), where galaxy properties were measured from the *Revised IRAS-FSC Redshift Catalogue* (RIFSCz; Wang et al. 2014). 4570 galaxies are included in both subsamples (HECATE-SDSS and HECATE-IRAS) and we discuss these further in Section 2.3. We applied additional cuts based on distances, 50–200 Mpc, the availability of reliable SFR and  $M_*$  measurements, and screening against AGN, which are discussed below. In the next few subsections, we detail the sample selection steps and record the changes to the sample size with each step in Table 1. Our final sample contains 60 212 unique (54 736 and 5 476 from HECATE-SDSS and HECATE-IRAS, respectively) galaxies. Fig. 1 shows their sky distribution.

## 2.1 The HECATE-SDSS subsample

The GSWLC catalogue is built primarily from spectroscopic SDSS data of the low-redshift ( $z < 0.3$ ) universe and includes galaxies that fall within the intersection of the *GALEX* + SDSS footprints, even if they are not detected in the ultraviolet. The resulting catalogue is optically selected and contains 730 288 unique galaxies with  $z = 0.01$ – $0.30$  and Petrosian  $r$ -band SDSS magnitude,  $r_{\text{petro}} < 18.0$ . Imposing the HECATE-SDSS Virgo-infall corrected radial velocity cut of  $14\,000 \text{ km s}^{-1}$  reduces the sample by a factor of 10. The GSWLC catalogue uses ultraviolet to optical photometry to fit the spectral energy distributions (SED), using the Bayesian approach

outlined in detail in Salim et al. (2016), and measure SFRs and  $M_*$  for the galaxies.

## 2.2 The HECATE-IRAS subsample

The HECATE catalogue also contains galaxies from the *Revised IRAS-FSC Redshift Catalogue*, which is nearly an all-sky ( $|b| > 20^\circ$ , i.e. excluding regions near the Galactic plane) catalogue of galaxies selected at  $60 \mu\text{m}$  combined with multiwavelength data from WISE All-sky Data Release (Wright et al. 2010), SDSS-DR 10 (Ahn et al. 2014), *GALEX* All-Sky Survey Source Catalogue (Seibert et al. 2012), 2MASS Redshift Survey (Huchra et al. 2012), and *Planck* Catalogue of compact sources (Planck Collaboration XVI 2014). The analysis by Kovlakas et al. (in preparation) allowed for a 10 per cent relative difference in the redshifts between the photometric redshifts from Wang et al. (2014) and the HECATE redshifts, and reconciled discrepant sources to create the HECATE-IRAS sample.

The HECATE-IRAS sample complements the HECATE-SDSS sample since the former has better all-sky coverage, but tends to be biased towards brighter (i.e. closer and larger) galaxies, which is evident from Fig. 2 and discussed further in Section 2.4, where we introduced a distance criterion to restrict our study to spatially unresolved galaxies. Therefore, our analysis has missing contributions from nearby galaxies, including metal-poor dwarf galaxies, which are discussed in Section 4.

## 2.3 Subsample overlap

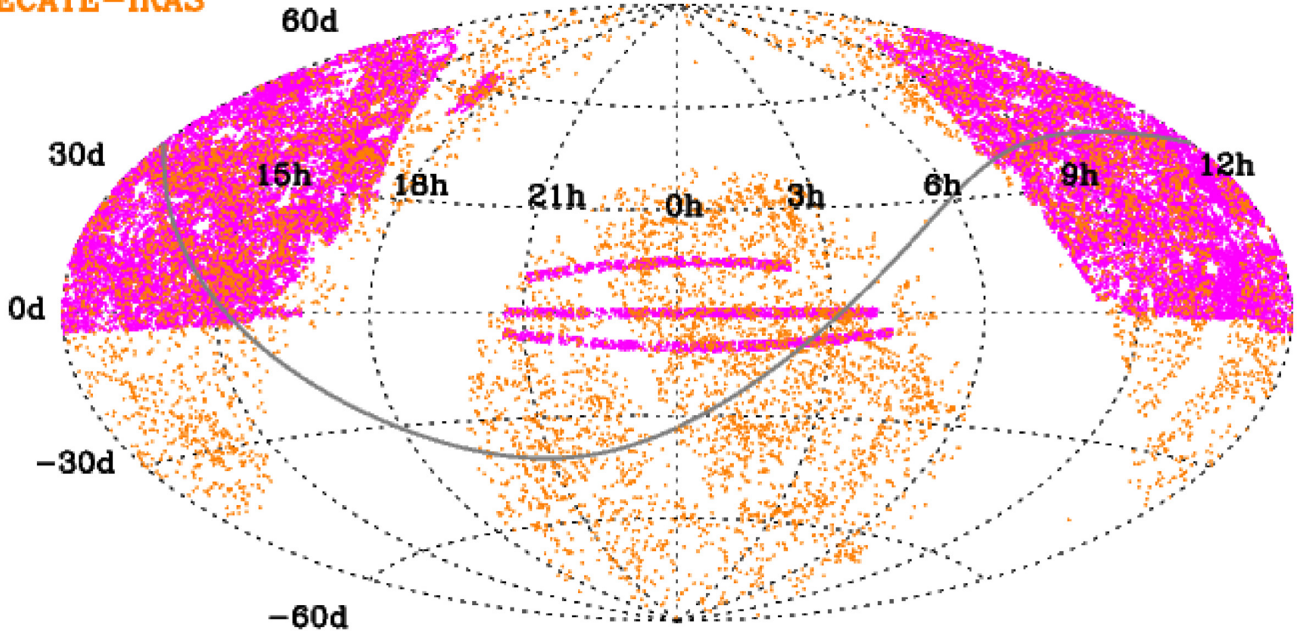
Between the HECATE-SDSS and HECATE-IRAS samples, 4570 galaxies are present in both samples, 97 per cent of these have distances between 50 and 200 Mpc. These duplicates were removed from the HECATE-IRAS subsample. However, we discuss these galaxies further in Section 2.5.2 since they provided the opportunity for additional consistency checks.

## 2.4 Selecting spatially unresolved galaxies

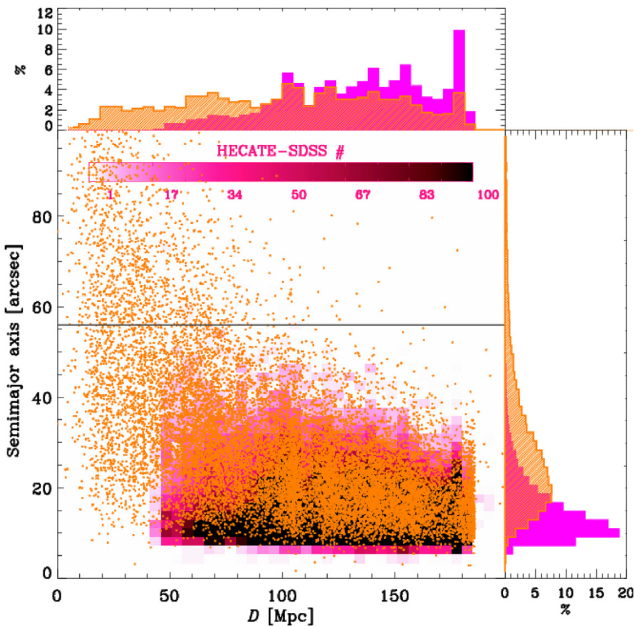
In this paper, we focused on galaxies outside the immediate universe (Local Group and vicinity) for two major reasons. The first was to concentrate on the cosmic volume over which we expect



HECATE-SDSS  
HECATE-IRAS



**Figure 1.** Distribution of sources that have well-measured star formation rates (SFRs) and stellar masses as well as reliable redshifts from the HECATE SDSS (magenta) and IRAS RIFSCz (orange) catalogues. The sky map is in equatorial coordinates and centred at  $\alpha = 0$  h and  $\delta = 0^\circ$ , with longitudinal and latitudinal gridlines marking every 3 h of right ascension and  $30^\circ$  of declination. The grey line marks the separation between the *eROSITA*-RU (above line; galactic longitude  $< 180^\circ$ ) and *eROSITA*-DE (below line; galactic longitude =  $180$ – $360^\circ$ ) skies.

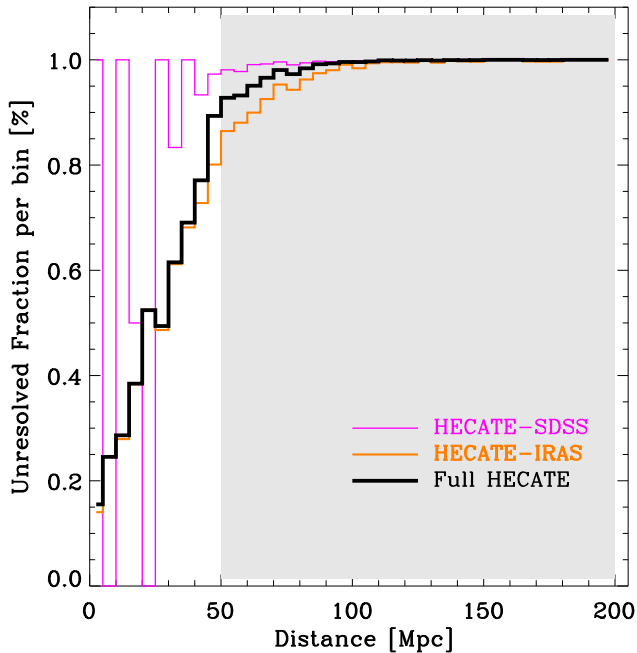


**Figure 2.** The central plot shows the distances and semimajor axis sizes of the galaxies in the parent samples. We show the full sample of HECATE-SDSS galaxies with the magenta colour, where the colour bar describes the number scaling. The orange points mark the HECATE-IRAS sample. The histograms at right and top show the size and distance distributions, respectively, shown for the HECATE-SDSS (magenta) and HECATE-IRAS (orange) samples. The solid black line marks twice the *eROSITA*HEW (56 arcsec). We expect galaxies smaller than this size to appear unresolved by *eROSITA*.

*eROSITA* surveys will break new ground. Pointed observations by observatories such as *XMM-Newton* and *Chandra* have been able to conduct relatively complete surveys within 20 Mpc of relatively complete galaxy samples assembled at optical/IR/UV wavelengths. We thus expect the great gains for *eROSITA*, as an all-sky survey, to be in the more distant universe, where it becomes impractical to target galaxies individually for follow-up. Secondly, for this paper we focused on spatially unresolved galaxies, since the X-ray scaling relations that we applied (described in more detail in Section 3) are based upon global (integrated) emission from galaxies. In Fig. 2, we show in magenta colour the distribution of distances and sizes (given by the semimajor axis in arcsec) for the HECATE-SDSS sample. The solid line marks the value for  $2 \times$  the *eROSITA*HEW (56 arcsec; Merloni et al. 2012), below which the galaxies are assumed to appear unresolved within *eROSITA*. Fig. 3 shows the fraction of unresolved galaxies as a function of distance for both subsamples (HECATE-SDSS in magenta and HECATE-IRAS in orange) and the total sample (black curve). Aiming to include the largest sample of unresolved galaxies, we find that the fraction of unresolved galaxies for the final sample (black curve) flattens out between 50 and 200 Mpc. Therefore, we chose to restrict our sample selection to these distances. Therefore, the total sample includes 81 805 galaxies. Of these, 4460 galaxies overlap, resulting in 77 345 unique galaxies with distances of 50–200 Mpc (see rows 2 and 3 in Table 1).

## 2.5 Measuring galaxy global properties

The HECATE catalogue incorporates derived galaxy properties (e.g. SFRs and  $M_\star$ ) from Salim et al. (2016) and Wang et al. (2014) for the SDSS and IRAS subsamples, respectively. However, since we strove for a consistent method of measuring SFRs and  $M_\star$  across the two subsamples, we independently measured these properties by



**Figure 3.** Fraction of galaxies that would appear unresolved by *eROSITA* as a function of distance. In this analysis, we assume that a galaxy is unresolved by *eROSITA* if its size is smaller than twice the *eROSITA* HEW, i.e.  $< 56$  arcsec. The magenta and orange lines distinguish the HECATE-SDSS and HECATE-IRAS subsamples, respectively, and the black thick line shows the total sample. We restrict our final sample to distances  $> 50$  Mpc where more than 90 per cent of the sources are unresolved. This is also the volume that has been relatively unexplored for X-ray emission from normal galaxies in surveys.

cross-matching our sample with the *GALEX* All Sky Survey (Bianchi et al. 2011), SDSS Data Release 12 (DR12; Alam et al. 2015), *AllWISE* All Sky Survey (Cutri et al. 2014), and 2MASS (Cutri et al. 2003) using TOPCAT (Taylor 2011). We describe the SFR and  $M_*$  measurement methods below. Note, however, 17 per cent of the galaxies do not have reliable SFRs because of missing *GALEX*NUV or *AllWISE*W4 magnitudes. In addition, 568 galaxies do not have available observations to allow any method for determining  $M_*$  (150 of these were missing SFRs as well). In order to retain the largest possible sample, we applied other methods to attain SFRs and  $M_*$ , which are also described in more detail below.

We show the distance, SFR, and  $M_*$  distributions for these subsamples in magenta (HECATE-SDSS) and orange (HECATE-IRAS) in Fig. 4 panels (a) to (c). We note that, based on the completeness discussed for the HECATE-SDSS (see Strauss et al. 2002) and the HECATE-IRAS (Wang et al. 2014) samples, we expect that the samples together cover most of the galaxies in the nearby Universe ( $D = 50$ –200 Mpc) possibly with the exception of ‘missing’ galaxies pertaining to extremely faint low surface brightness, ultradiffuse, and/or dwarf galaxies, which are unlikely to be detected by *eROSITA* due to their expected low SFR and  $M_*$ . Therefore, we expect that the distribution of SFR and  $M_*$  displayed in the middle and right-hand panels of Fig. 4 effectively represent the majority of the detectable local population of galaxies.

### 2.5.1 Star formation rates

The SFRs are based on the UV+IR to account for stars forming both outside of and within dust-obscured regions. We applied the

conversions from Bell et al. (2005):

$$\text{SFR} (M_{\odot} \text{ yr}^{-1}) = 9.8 \times 10^{-11} (L_{\text{IR}} + 3.3L_{\text{UV}}), \quad (1)$$

where  $L_{\text{IR}}$  and  $L_{\text{UV}}$  correspond to the total IR (8–1000  $\mu\text{m}$ ) and UV ( $\nu L_{\nu}$  at 2800  $\text{\AA}$ ) luminosities, respectively, in units of solar luminosity. SFRs were calculated using the  $L_{\text{UV}}$  determined from *GALEX* data,<sup>4</sup> the  $L_{\text{IR}}$  determined from WISE data,<sup>5</sup> and equation (1). The factor of 3.3 is the typical correction factor that brings  $\nu L_{\nu}$  (at 2800  $\text{\AA}$ ) to the integrated  $L_{\text{UV}}$ . We converted monochromatic 22  $\mu\text{m}$  luminosity, based on the WISE channel 4 photometry, to  $L_{\text{IR}}$  using the Chary & Elbaz (2001) IR SED template.

Among the  $D = 50$ –200 Mpc HECATE-SDSS galaxies, the method described above allows SFR measurements for  $\sim 71$  per cent of the sample. For the rest, the GSWLC catalogue provides two alternate SFR estimates (see Salim et al. 2016 for more details): (1) based on SED-fitting, (2) based on the IR from ‘unWISE’ (Lang, Hogg & Schlegel 2016) where SDSS detections serve as forced photometry priors.

Comparing our preferred method (equation 1), applied to the majority of galaxies, with the ‘unWISE’ SFRs, we find excellent agreement, with median  $\Delta \log(\text{SFR}) = 0.03 \text{ dex}$  [ $M_{\odot} \text{ yr}^{-1}$ ] and  $\sigma(\Delta \log(\text{SFR})) = 0.2$  [ $M_{\odot} \text{ yr}^{-1}$ ]. Therefore, for another 7 701 HECATE-SDSS galaxies, we adopted the SFRs given by the ‘unWISE’ method in the GSWLC catalogue. SFRs are also available in the GSWLC catalogue by SED fitting the UV to optical data. However, by comparing this method with the others described above, we found significant differences, especially at  $\log \text{SFR} < 0$ , where the other methods that included the infrared measured SFRs were 1–3 orders of magnitude higher than those from the GSWLC catalogue. Therefore, we eliminated these from our analysis.

For the HECATE-IRAS sample, nearly half of the galaxies have SFRs measured using equation (1). Many of the remaining galaxies are missing SFRs due to missing UV and IR data. Therefore, our analysis includes 61 067 galaxies between 50 and 200 Mpc with valid SFRs (see rows 4–6 in Table 1).

### 2.5.2 Stellar masses

To derive the stellar masses, we used the following relation from Zibetti, Charlot & Rix (2009):

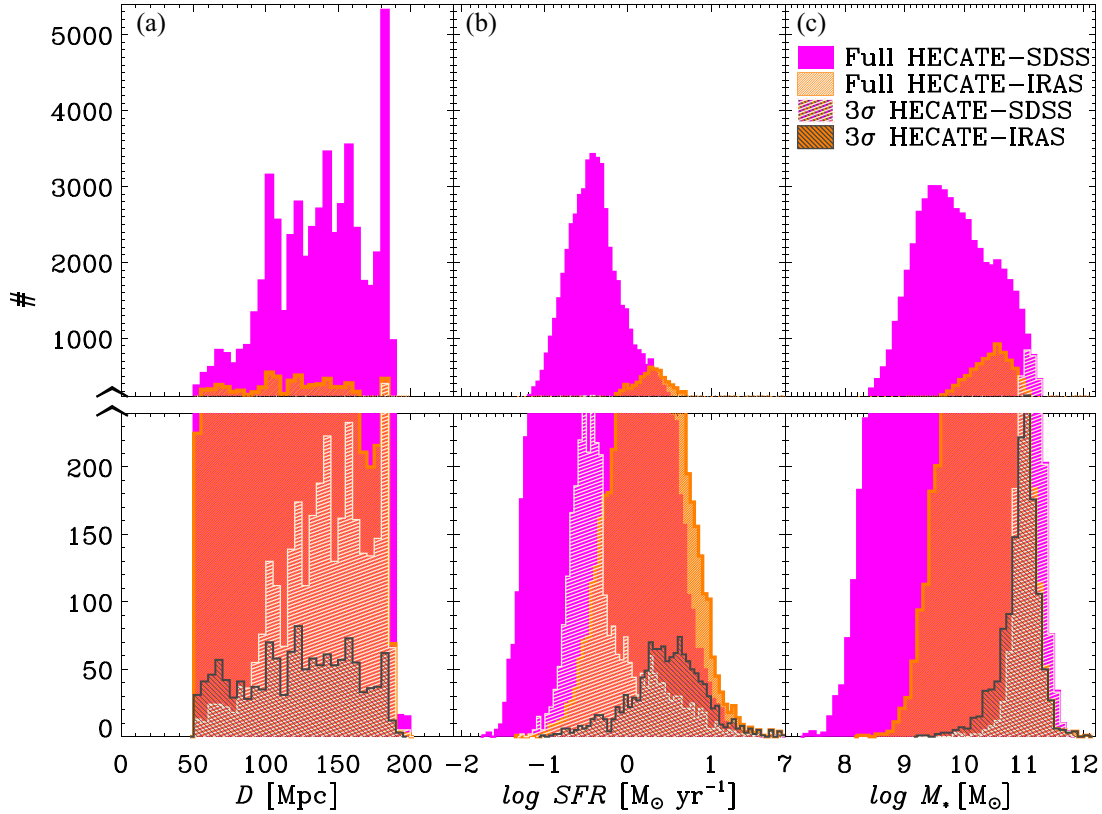
$$\log M_* = \log L_K + 1.119 * (g - r) - 1.257, \quad (2)$$

where  $g - r$  is the colour in SDSS  $g$  and  $r$  bands and  $M_*$  and  $L_K$  are in solar units. Unfortunately,  $L_K$  and  $g - r$  colours are only available for some galaxies in our subsamples. We describe the stellar mass measurements in more detail in the next few subsections, as their treatment depended upon the subsample to which they belong.

For the HECATE-SDSS galaxies,  $g - r$  colours are available for all the galaxies, however  $K$ -band luminosities are only available for 69 per cent of the sample, therefore, the  $M_*$  measurement either followed the methods outlined above (using equation 2) or used the stellar mass derived by Salim et al. (2016), based on SED fits to the UV (*GALEX*) and optical (SDSS) data. Comparing the masses determined by the two methods on the 69 per cent with  $K$ -band

<sup>4</sup>We applied extinction corrections to the NUV magnitudes as given by Salim et al. (2016).

<sup>5</sup>Details describing the WISE All-Sky catalogue and 22  $\mu\text{m}$  flux measurements can be found at <http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/index.html> and in Wright et al. (2010), Jarrett et al. (2011), and Cutri et al. (2014).



**Figure 4.** Histograms for distance (a), SFR (b), and stellar mass (c) for the different samples: full HECATE-SDSS sample (magenta), full HECATE-IRAS sample (orange), and expected  $3\sigma$  detections in the *eROSITA* 4-year survey from the HECATE-SDSS (shaded white) and HECATE-IRAS samples (shaded dark grey) based on the analytic simulation of pure galaxies (Sections 3.1 – 3.3, i.e. no AGN components).

observations, we find the two methods agree very well, with average difference of  $\Delta \log(M_*/M_\odot) < 0.08$  dex, with no observed bias with  $M_*$ .

Therefore, our HECATE-SDSS subsample includes 66 849 galaxies, with the breakdown between the two methods given in column 4 of rows 7–10 of Table 1.

For the HECATE-IRAS subsample, we faced a challenge in measuring the stellar masses because  $g-r$  colours are not available for 66 per cent of the galaxies. The majority of those with  $g-r$  colours are galaxies which are in both subsamples (i.e. also in HECATE-SDSS, discussed in Section 2.3) and we used the average  $g-r$  colours from this subsample to measure  $M_*$  for the HECATE-IRAS galaxies that are missing observed  $g-r$  colours.

Fig. 5 (left-hand panel) shows the distribution of  $g-r$  colour in the overlapping sample based on the SDSS photometry. We estimated  $M_*$  based on the median  $g-r = 0.7$  and show the distribution of ratios of the estimates compared to stellar masses measured using the observed  $g-r$  colours in the right-hand panel of Fig. 5. Based on this distribution, we estimated that our method introduces uncertainties in  $M_*$ ,  $\sigma \log(M_*) \approx 0.18$  dex in the HECATE-IRAS sample, shown by the grey line. The uncertainties from this method on  $\lesssim 10$  per cent of the total sample is less than the scatter in the X-ray- $M_*$  scaling relation (given below in equation 5), therefore, applying the method introduces insignificant errors on the final results.

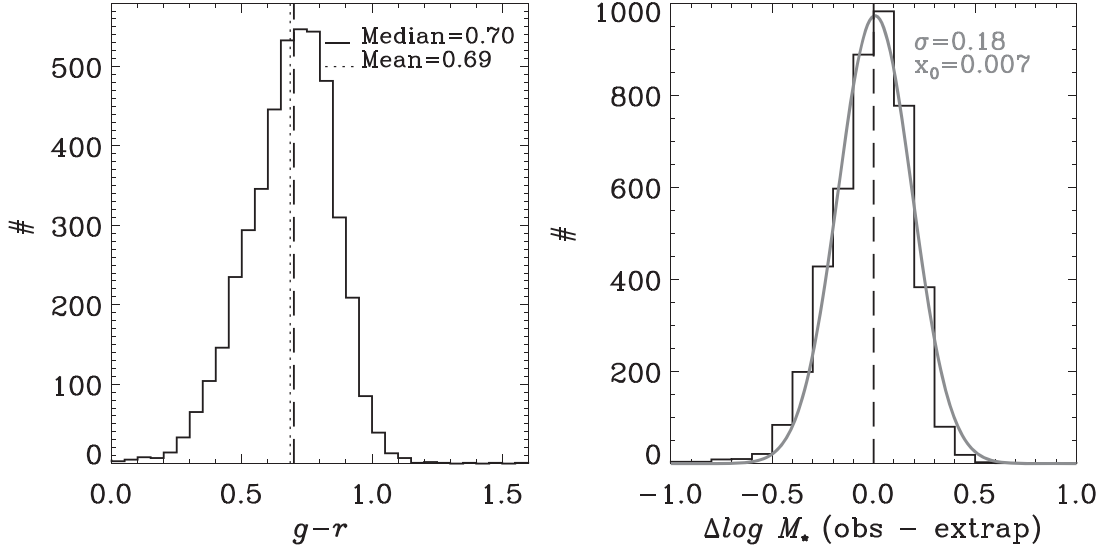
While the overlapping galaxies were eliminated from the HECATE-IRAS sample, they remained in the HECATE-SDSS sample and, therefore, we measured  $M_*$  for these using their observed (true)  $g-r$  values.

Table 1 presents the numbers of galaxies where we measured stellar masses using either observed (row 7) or extrapolated (row 9)  $g-r$  colours, or via SED-fitting (row 8). Of these, 79 per cent also have valid SFRs (row 9). Any inconsistency between the different SFR and  $M_*$  methods has no impact on the final results of this study.

## 2.6 AGN screening

Since the goal of this paper is to study the X-ray emission from normal galaxies, we also want to screen against AGN. In addition, the SFR and  $M_*$  estimates are uncertain if AGN are present, since the IR fluxes could be due to AGN emission rather than from star formation activity, i.e. the assumed relation from equation (1) would not apply. Therefore, for the HECATE-SDSS galaxies, we screened against AGN by selecting the galaxies with ‘flag\_sed’ = 0 based on the SED-fitting flag from GSWLC (Salim et al. 2016), which allowed us to select galaxies that are robustly fit by galaxy spectral templates. In contrast, the catalogue identifies broad-line spectra typical of Type 1 AGN and Seyferts and those spectra with poor fits, with flags set to 1 and 2. Admittedly, screening against the broad-line AGN via this method does not remove narrow-line AGN, which may be the majority of AGN. The HECATE-IRAS galaxies do not have any reliable AGN screening diagnostics available. However, we address potential AGN contamination in more detail in Section 3.5. Therefore, our final sample includes 60 212 galaxies at distances of 50–200 Mpc (screened against AGN) with reliable SFRs and stellar masses.





**Figure 5.** For the galaxies in the HECATE-IRAS subsample that lack observed  $g - r$  colours, we estimated the  $M_*$  based on the 4460 galaxies found in both HECATE-SDSS and HECATE-IRAS sample. Left-hand panel: Colour distribution for the overlapping sample. The mean (0.69) and median (0.70) values are very similar, and we adopted the median  $g - r$  colour = 0.70 for the remaining galaxies in the HECATE-IRAS sample. In our final  $M_*$  measurements, we used the measured colours where available (i.e. for the overlapping sample) and only used the extrapolated colours for those in the HECATE-IRAS sample with missing  $g - r$  colours (see discussion in Section 2.3). Right-hand panel: Comparison of the extrapolated mass measurements with those using the observed  $g - r$ . The grey line shows the fit to the distribution with a Gaussian that is centred at  $\approx 0$  with  $\sigma = 0.18$  dex in the  $\log M_*$  values.

### 3 SIMULATING THE NORMAL GALAXY POPULATION

For our analysis, we assumed two sources of X-ray emission within normal galaxies: the first from XRBs and the second from the hot ISM component. Hereafter, we refer to the ISM component as the ‘gas’ component. The total X-ray luminosity for the galaxy ( $L_{X, \text{gal}}$ ) is the sum of the XRB and gas components,

$$L_{X, \text{gal}} = L_{X, \text{XRB}} + L_{X, \text{gas}}. \quad (3)$$

The applied scaling relations are sufficient for the *eROSITA* predictions, however understanding the complexities and characterizing the nature of galaxy-to-galaxy scatter in these relations are major goals for X-ray studies of normal galaxies. In particular, in-depth studies of nearby galaxies show that these scaling relations may vary with galaxy properties (e.g. stellar age, metallicity). In this paper, we use the zeroth-order estimates in the scaling relations to predict *eROSITA* detections of normal galaxies.

#### 3.1 Scaling relations and spectral parameters

To estimate the X-ray luminosity that originates from XRBs, we used the scaling relation from Lehmer et al. (2016, their equation 15), where the X-ray luminosity from XRBs was further separated into the low-mass XRB (LMXB) and high-mass XRB (HMXB) components. The former scales with stellar mass ( $M_*$ ) and the latter with star formation rate (SFR), as given by

$$L_{2-10, \text{XRB}}(z) = L_X(\text{LMXB})(z) + L_X(\text{HMXB})(z) \quad (4)$$

$$= \alpha_0(1+z)^\gamma M_* + \beta_0(1+z)^\delta \text{SFR}, \quad (5)$$

where  $\log \alpha_0 = 29.30 \pm 0.28$ ,  $\log \beta_0 = 39.40 \pm 0.08$ ,  $\gamma = 2.19 \pm 0.99$ ,  $\delta = 1.02 \pm 0.22$ . For our analysis,  $z = 0$  and we applied this equation using the stellar masses and SFRs discussed above to predict the X-ray luminosity due to XRBs. The XRB luminosity from

the above equation is for the 2–10 keV band. We describe how we combine the XRB and hot gas X-ray luminosity given different bands into output 0.5–10 keV fluxes and counts, using additional spectral models, below.

The hot gas component of galaxies differs for mass-dominated versus star-forming galaxies, with the former having hot gas scaling with the galaxy’s mass, reflecting the overall size of the reservoir potential for hot ISM and the latter scaling with current star formation activity. To calculate the hot gas component, we first separated our sample into mass-dominated and star-forming galaxies, using the specific star formation rate:  $\text{sSFR} = \text{SFR}/M_*$ . Galaxies dominated by star formation ( $\log \text{sSFR} > -10.1 \text{ yr}^{-1}$ ) were considered star-forming and the lower sSFR galaxies were mass-dominated in our analysis.<sup>6</sup> In Fig. 6, we show the distribution of sSFR for the two HECATE parent samples, SDSS (magenta) and IRAS (orange). For the hot gas in star-forming galaxies, we used the relation given in Mineo, Gilfanov & Sunyaev (2012b), which provides the 0.5–2 keV luminosity and was converted to our chosen IMF (from Salpeter to Kroupa):

$$L_{0.5-2, \text{gas}}/\text{SFR} = (12.4 \pm 0.2) \times 10^{38} \text{ erg s}^{-1} \quad (6)$$

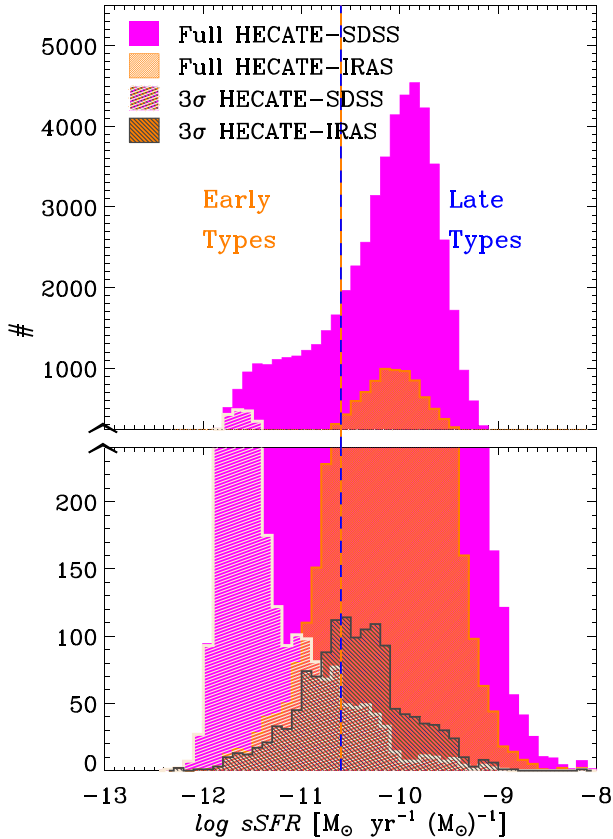
For mass-dominated galaxies ( $\log \text{sSFR} < -10.1 \text{ yr}^{-1}$ ), we used the equation from Kim & Fabbiano (2015), which estimates the hot gas contribution to the 0.3–8 keV band as follows:

$$\log(L_{0.3-8, \text{gas}}/10^{40} \text{ erg s}^{-1}) = A \log(L_K/10^{11} L_{K\odot}) + B \quad (7)$$

and applied their coefficients for the ‘Full’ sample:  $A = 2.98 \pm 0.36$  and  $B = -0.25 \pm 0.11$ . Since the Kim & Fabbiano (2015) equation is expressed in terms of  $L_K$ , we converted this hot gas relation to stellar

<sup>6</sup>Later in the analysis, we discuss late- and early-type galaxies, divided by  $\log \text{sSFR} = -10.6 \text{ yr}^{-1}$ , which is a more moderate limit that better divides the galaxy types (see Fig. 6)





**Figure 6.** Distribution of sSFR, with the same sample representation as in Fig. 4. The bimodal distribution of galaxies is apparent for the HECATE-SDSS galaxies (magenta), where  $\log \text{sSFR} = -10.6$  (dashed line) falls between the two peaks and separates the galaxy populations into early-types ( $< -10.6$ ) and late-types ( $> -10.6$ ).

mass, by using the relation given by Lehmer et al. (2014),

$$M_*/L_K = 0.66 M_\odot / L_{K,\odot} \quad (8)$$

For the most accurate determination of detectability by *eROSITA*, we modelled X-ray spectra as well as the total  $L_X$  of the sources. To model the X-ray spectra in these galaxies, we used a  $\Gamma = 1.8$  power law for the XRBs and an APEC model for the gas, multiplied by a PHABS model to account for Galactic absorption, where  $N_H$  was estimated from the sky position and the `ftools` ‘nh’ tool. We used the default abundance model given by Anders & Grevesse (1989). To estimate the temperature of the gas,  $kT$ , from the X-ray luminosity predicted above, we applied the equation from Kim & Fabbiano (2015) for their ‘Full’ sample:

$$\log(L_{X,\text{gas}}/10^{40} \text{ erg s}^{-1}) = A \log(kT_{\text{gas}}/0.5 \text{ keV}) + B, \quad (9)$$

where  $A = 5.39 \pm 0.60$  and  $B = 0.16 \pm 0.43$ .

For star-forming galaxies, Mineo et al. (2012b) fit the spectra of hot gas in galaxies with `mekal` models with  $kT = 0.24 - 0.71$  keV, with the same abundance model. Therefore, for our sample of galaxies that have  $\log \text{sSFR} > -10.1 \text{ yr}^{-1}$ , we followed a similar procedure as described above, but replaced the APEC model with `mekal` and used  $kT = 0.5$  keV as a baseline value. In Section 3.3, we uniformly sampled the  $0.24 - 0.71$  keV  $kT$  range to account for the observed range of temperature values for the hot gas component of star-forming galaxies.

Using the distances given in the HECATE catalogue, we converted our luminosities into predicted fluxes for the galaxies. Additionally, we accounted for the encircled energy fraction of the PSF by multiplying the input fluxes by 50 per cent to simulate the expected flux within the PSF. These fluxes for the components in the specified bands (i.e. 2–10 keV for the XRB) set the normalization for the model components. The final flux for each galaxy refers to the model flux output from the `XSPEC` model, which we output for the standard *eROSITA* bands (0.2–0.5, 0.5–1, 1–2, 2–4.5, 4.5–10, and 0.5–10 keV). However, for simplification, henceforth our results focus on the 0.5–10 keV band, unless specified.

### 3.2 Determining *eROSITA* detections using response files

We set the spectral normalization for the different spectral components (in the specified bands) using the fluxes calculated above for the gas and XRBs, as discussed in the previous section. We used the on-axis redistribution matrix file (RMF) file for all valid patterns<sup>7</sup> and created a combined ancillary response file (ARF) file for all seven telescopes. The ARF file was based on the 200 nm Al + 200 nm PI filter configuration for telescopes 1, 2, 3, 4, and 6 and the 100 nm Al + 200 nm PI filter for telescopes 5 and 7.<sup>8</sup> Using the model spectra with these files, we predicted the count rates for the 0.5–2 keV and 2–10 keV bands. We combined these counts to report 0.5–10 keV results throughout. We converted these into the expected counts for each galaxy in the *eROSITA* all-sky survey, using the vignetting-corrected, position-dependent 4-yr exposure map produced by the Simulation of X-ray Telescopes (SIXTE) `exposure_map` tool based on the attitude file available from the *eROSITA* instruments installation for SIXTE. This attitude file is representative for the survey, although small deviations between this attitude file and the survey are expected due to the prolonged performance verification phase of the mission.

To calculate the number of background counts expected in our observations, we used the model background values given for the different *eROSITA* bands,<sup>9</sup> given in units of counts  $\text{s}^{-1} \text{ arcmin}^{-2}$ . Therefore, to get the total number of background counts for each source, we multiplied by the same exposure used above and by the area of the galaxy. Since our galaxies are unresolved, we used the PSF area derived from the 28 arcsec HEW for the  $\sim 80$  per cent with sizes smaller than the PSF area. For  $\sim 13\,000$  galaxies with larger sizes, we used the area of the galaxy, based on the available semiminor and semimajor sizes.

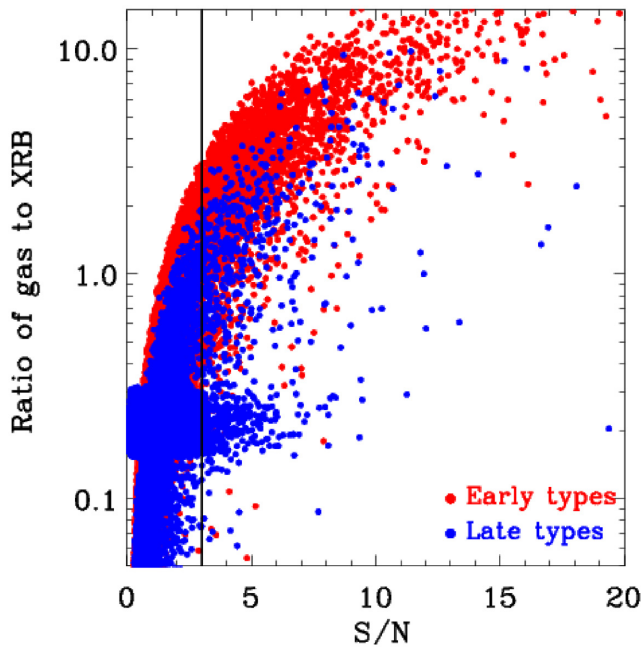
We predicted the signal-to-noise ratio (SNR) as  $\sigma = S/\sqrt{S+B}$  based on the expected source (S) and background (B) counts.<sup>10</sup> Fig. 7 shows the ratio of gas to XRB component in the 0.5–10 keV flux as a function of the 0.5–10 keV SNR in the two galaxy populations, e.g. early types (red) versus late types (blue), with the vertical line marking the  $3\sigma$  detection limit. We find that the detected galaxies are dominated by early-types, due to their higher ratio of gas to XRB emission. The cluster of late-type galaxies with gas to XRB ratios between 0.15 and 0.3 refer to the extreme late-type population with

<sup>7</sup>The RMF file, `rmf01_sdtq.fits`, is available from <https://wiki.mpe.mpg.de/eRosita/erocalib/calibration>.

<sup>8</sup>The ARF files, `arf01_100nmAl_200nmPI_sdtq.fits` and `arf01_200nmAl_200nmPI_sdtq.fits`, are available from <https://wiki.mpe.mpg.de/eRosita/erocalib/calibration>.

<sup>9</sup>The *eROSITA* background model is available at [http://www2011.mpe.mpg.de/erisita/eROSITA\\_background\\_v8.pdf](http://www2011.mpe.mpg.de/erisita/eROSITA_background_v8.pdf)

<sup>10</sup>As an additional test, we also calculated the Poisson S/N using equation 7 in Li & Ma (1983), but found that this only made a 2 per cent difference in the number of  $3\sigma$  detections.



**Figure 7.** SNR versus ratio of gas to XRB component in the expected 0.5–10 keV source counts. The vertical black line marks the  $3\sigma$  detection limit and the red and blue points mark the early- and late-type galaxies.

$\log sSFR > -10.1 \text{ yr}^{-1}$ , whose prescription for the hot gas emission (see equation 6) is different from the other galaxies (see equation 7).

In the following sections, we included the effects of scatter in the scaling relations presented in this section and discuss additional caveats to *eROSITA* ‘detections’. We show the distance, SFR,  $M_*$ , and  $sSFR$  distributions for the detected galaxies in Figs 4 and 6 for the SDSS (white/magenta striped) and IRAS (grey/orange diagonally striped) subsamples.

### 3.3 Effects of scatter on ‘detectability’

Thus far, we have presented a simplified view of X-ray emission from normal galaxies, assuming that  $L_{X, \text{gal}}$  follows the scaling relations. In practice, galaxy-to-galaxy variations from these scaling relations can be quite large, and have been shown to vary systematically with second-order galaxy properties such as metallicity and stellar age (e.g. Basu-Zych et al. 2013a,b; Douna et al. 2015; Brorby et al. 2016; Lehmer et al. 2016, 2017, 2019, and references therein). In this section, we explore the effect of scatter from the above scaling relations on our predictions of whether a galaxy can be detected.

We performed Monte Carlo simulations, by creating 500 realizations of the XRB and gas luminosities based on the uncertainties given in the scaling relations. Specifically, each realization draws randomly from a Gaussian distribution (with the exception of the late-type temperature,  $kT$ , which was uniformly drawn within the 0.24–0.71 keV range; Mineo et al. 2012b) that is centred on the values given in Section 3 with  $1\sigma$  standard deviations matching the given uncertainties.

Therefore, each galaxy has a distribution of signal-to-noise values based on the combination of scatter within the different relations that contribute to the integrated X-ray emission from the galaxy.

In the next section, we explore how well these predicted signal-to-noise measurements match ‘detections’ within *eROSITA* all-sky simulated data.

### 3.4 *eROSITA* SIXTE simulations

As an alternative test of the results for the *eROSITA* 4-yr survey (eRASS:8) simulations completed in the first part of Section 3, we produced end-to-end simulations using the SIXTE software<sup>11</sup> (Dauser et al. 2019). SIXTE uses an instrument model and astrophysical source models to create events files that can be used for analysis, such as image generation and source detection. It uses a Monte Carlo based approach based on simulating single photons. These are imaged on to the detector using the lab-measured energy-dependent vignetting and PSF function to correctly model the imaging of each source as it passes through the field of view. While such a detailed simulation is very computationally expensive, it is the most correct description of the complex *eROSITA* instrument in a slew survey and therefore serves as a perfect test to verify the previously obtained results.

SIXTE takes FITS-based Simulation inPUT (SIMPOT) files as inputs to the simulator, which for our purposes required the source position ( $\alpha, \delta$ ), source flux, source spectral shape, and for various background components the source image. We followed the same procedures outlined in Section 3.1 to determine galaxy X-ray fluxes and spectra to create SIMPOT files for each of the 74 182 galaxies in the sample. These individual galaxy SIMPOTs served as the ‘source’ component of the simulation. The background component was comprised of five different SIMPOTs<sup>12</sup> that were used by Dauser et al. (2019, please refer to this paper for details of SIMPOT creation) to simulate the 6-month *eROSITA* survey (eRASS:1):

- (i) Catalogue of  $10^6$  AGN with unique source spectra (and references therein; Cappelluti et al. 2007; Dauser et al. 2019)
- (ii) Catalogue of  $3 \times 10^6$  galaxy clusters with images and unique spectra (Finoguenov et al. 2015)
- (iii) Catalogue of 125 000 *ROSAT* all-sky survey point sources (Voges et al. 1999)
- (iv) The *ROSAT* all-sky survey soft X-ray background (Snowden et al. 1997)
- (v) Galactic ridge emission (Türler et al. 2010)

In addition, SIXTE also incorporates an estimate of the particle (non-X-ray) background for *eROSITA*, based on a flat spectrum with the expected background rate. We also included this component in our simulations.

We simulated events for each galaxy using the input source (galaxy) and background (five separate components) SIMPOTs. To avoid long simulation times for each galaxy, we filtered the background SIMPOT files to only include background sources that were located within the *eROSITA* FOV for a given galaxy. We used the response files noted in Section 3.2 and the *eROSITA* 4-yr attitude file<sup>13,14</sup> that uses 10 s pointing intervals that allow taking into account vignetting. For each galaxy, SIXTE created an events file for each of the 7 telescopes, which were merged using the FTOOL `ftmerge` in preparation for source detection.

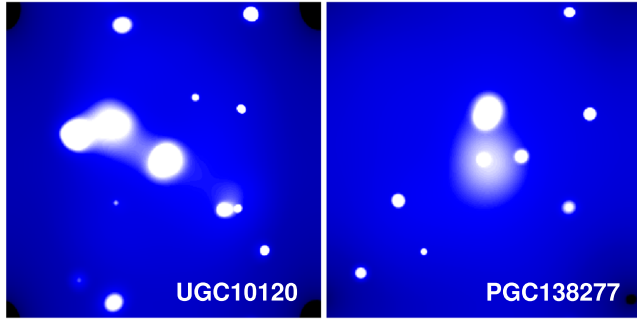
We created images of each galaxy in the 0.5–10, 0.5–2, and 2–10 keV energy bands from the output SIXTE event files using the SIXTE tool `imgev`. Source detection was completed using `wavdetect`

<sup>11</sup><https://www.sternwarte.uni-erlangen.de/research/sixte>

<sup>12</sup>Provided by Thomas Dauser and Philipp Weber

<sup>13</sup>eRASS\_Pc87M55\_3dobi\_att\_remeis.fits; available on the SIXTE website

<sup>14</sup>The attitude file used here is based on an earlier launch date than the real launch date, such that exposure times for individual sources will be slightly different from the real exposures. This, however, will not affect the overall statistical properties of our results.



**Figure 8.** *eROSITA* SIXTE simulation images (0.5–8 keV) of galaxies detected at high (left-hand panel) and low (right-hand panel) significance,  $54\sigma$  and  $3\sigma$ , centred on each image. Images are 1 degree on a side (*eROSITA* FOV) and have been smoothed using `csmooth` to bring out faint features (image intensity scales differ).

**Table 2.** SIXTE galaxy detections for eRASS:8. Detections were restricted to distances  $D \geq 50$  Mpc and  $\text{SNR} \geq 3$ .

Full	Soft	Hard	All
3069	3065	170	3318

with the scales parameter set to ‘1, 2, 4, 8, 16’ for the decreased spatial resolution per pixel of 9.65 arcsec compared to *Chandra*. We selected point sources that were detected within 15 arcsec of a galaxy’s coordinates, given that we do not expect a galaxy to be detected at a position beyond this value. In cases where more than one source was present within this radius (i.e. a background source), we selected the source with the smallest angular separation from the simulated galaxy position. We also cross-matched galaxy positions to each of the background point source catalogues within 15 arcsec, and checked to determine if a background source was mistaken for a galaxy detection.

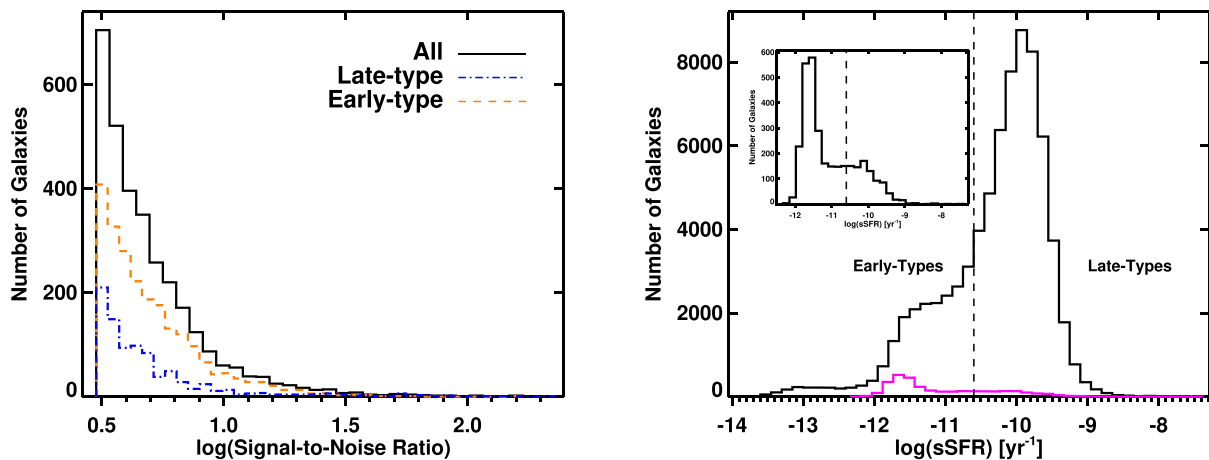
We selected galaxies with distances  $D \geq 50$  Mpc, motivated by the fact that we simulated each galaxy as a point source, and thus galaxies at this distance can be well-approximated as point sources (particularly within the average *eROSITA* survey-mode PSF HEW

of 28 arcsec within the FOV). Spatially resolved nearby galaxies require more detailed SIMPUT creation via inclusion of individual point sources and their spectra as well as images indicating the distribution of hot gas and the associated spectral components. This type of simulation is computationally intensive and beyond the scope of this paper, but has been completed for M31, M33, and NGC 6946 (N. Vulić). Based on Fig. 2, the majority ( $\sim 80$  per cent) of galaxies have angular sizes smaller than the *eROSITA* PSF HEW. We assume that the remaining galaxies have X-ray surface brightness profiles similar to their optical profiles, such that most of the X-ray emission is concentrated at the centre of the galaxy, and thus can be represented by a point source. In Fig. 8 we show simulated images from SIXTE for galaxies detected at  $54\sigma$  (left-hand panel) and  $3\sigma$  (right-hand panel).

In Table 2, we summarize the results of galaxy detections for each energy band, and the total number of unique galaxies detected across all energy bands. We calculated the SNR using the same method stated in Section 3.2. After inspecting the images we found that sources at a significance of  $3\sigma$  were clearly detected and could not be attributed to background fluctuations. In Fig. 9 we show histograms of the SNR and sSFR for all detections ( $\text{SNR} \geq 3$ ,  $D \geq 50$  Mpc). We also indicate the split between early-type and late-type galaxies, detecting 2 216 and 1 102, respectively.

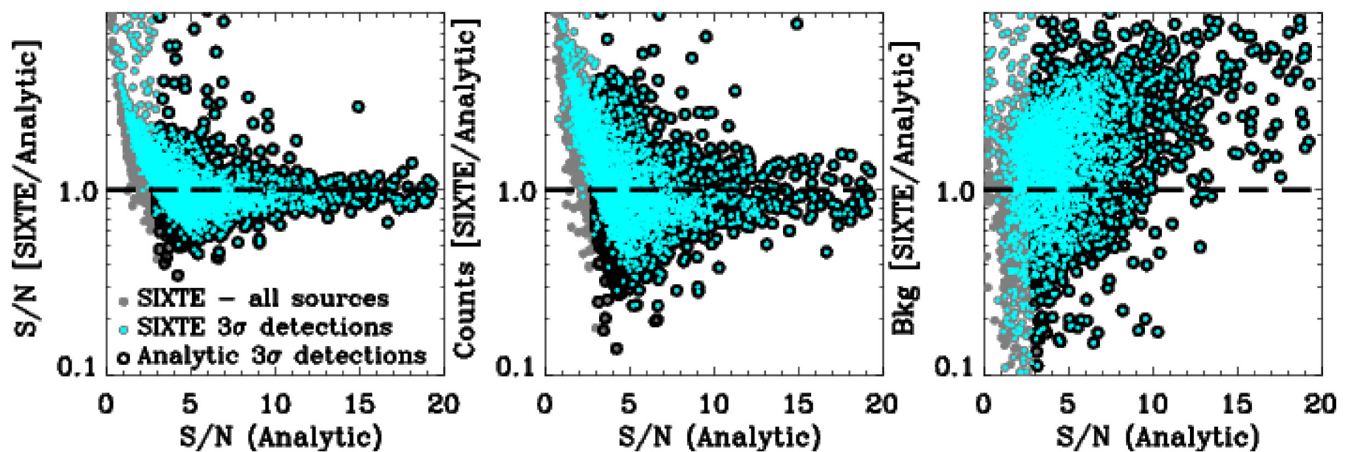
We compare the SIXTE results with those from the analytic method in Fig. 10, showing the ratios of SNR (left-hand panel), source counts (middle), and background counts (right-hand panel) as a function of the analytic SNR. Overall, the agreement is fairly good. The SIXTE method yields fewer  $3\sigma$  detections in the 0.5–10 keV band (3 072) than the analytic method (4 029), which is likely due to higher background counts in the SIXTE simulations (see right-hand panel of Fig. 10), which include additional components, e.g. clusters, extended emission from the Galactic ridge along the plane.

The goal of this section is to provide a test between two different methods for detecting galaxies with *eROSITA*. While SIXTE provides an accurate simulation of the *eROSITA* survey, it is also computationally intensive. The additional analyses (e.g. Sections 3.3 and 3.5) would take prohibitively long with SIXTE. Since there is overall good agreement between the results of our method, for the remainder of this paper we build on the analysis discussed in Sections 3.1–3.3 and present the results based on this analytic method.



**Figure 9.** *eROSITA* SIXTE simulation unique galaxy detections ( $\text{SNR} \geq 3$ ,  $D \geq 50$  Mpc) across all energy bands. Left-hand panel: SNR for all detections (solid black line) and split by sSFR into late-type (dot-dashed blue line) and early-type (dashed red line) galaxies. Right-hand panel: Histogram of sSFR for all galaxies in the parent sample (black, Fig. 6), and all unique detections (magenta). The inset shows all unique detections (magenta histogram). The majority (67 per cent) of detected galaxies are early-types, based on stellar mass being the dominant driver for X-ray luminosity from the scaling relations.





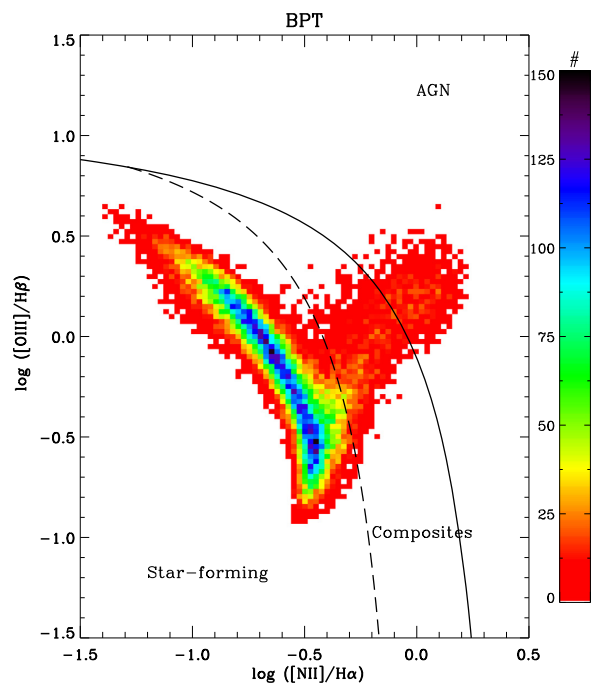
**Figure 10.** Ratio between SIXTE simulations and the analytic approach, described in Section 3, as a function of the analytic SNR for SNR (Left-hand panel), source counts (Middle), and background counts (Right-hand panel). The gray points mark all sources matched to input galaxies in the SIXTE simulation, with cyan points showing those detected at  $>3\sigma$ . The black open circles mark the  $>3\sigma$  detections in the analytic method.

### 3.5 AGN contamination

While every effort has been made to focus on the normal galaxy population, based on the optically selected galaxy catalogue, there is some possibility that some of these sources are in fact low-luminosity or obscured AGN. In many observed cases, emission from XRBs and low luminosity AGN are both present, and while it is often unclear which dominates the X-ray emission, both contributions are significant. Therefore, our galaxy prescription described above would likely underestimate the actual X-ray emission in such sources. However, predicting the X-ray emission from accreting supermassive black holes within galaxies is not straightforward and we present this section as an illustration of how the contribution from AGN might affect the observed distribution of *eROSITA*-detected sources.

We have excluded obvious broad-line AGN based on fits to the SED (Salim et al. 2016), as discussed previously in Section 2.6. However, in this section we statistically attempt to account for the X-ray emission from potential AGN. The dominant contribution from star formation versus AGN processes within galaxies is often determined via optical emission line ratios (Baldwin, Phillips & Terlevich 1981, e.g. BPT diagnostic), where star-forming galaxies are separated from AGN-dominated and ‘composite’ sources by a set of theoretical and empirical curves. However, in practice, these curves are porous in the sense that X-ray detected AGNs have been found within the ‘purely’ star-forming region (Agostino & Salim 2019; LaMassa et al. 2019). Jones et al. (2016) and Stampoulis et al. (2019) present methods to extract the AGN contribution fraction based on a source’s location on the BPT diagram. We note that these fractions are the contribution to the optical emission lines, not the X-ray emission.

We also caution that the objects classified by their optical emission lines may have a complicated mix of sources producing their X-ray emission. For example, several studies have established that LINERs are a heterogeneous class (see Molina et al. 2018). Eracleous et al. (2002) and Flohic et al. (2006) analysed the *Chandra* observations for several well-known, nearby LINERs and found that the X-ray emission, even within the nuclear region, are inconsistent with accreting SMBH and additional energetic sources may be responsible for powering the optical emission lines. Therefore our attempt to predict the X-ray emission from galaxies outside the ‘star-forming’ locus in Fig. 11 aims to provide a rough estimate for scaling the relative contributions from XRBs and AGN.

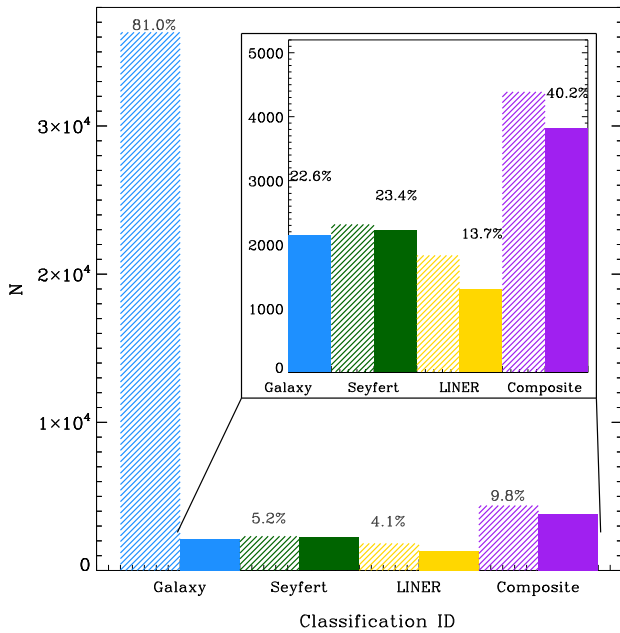


**Figure 11.** BPT diagram, which uses the  $\log[O\text{ iii}]/H\beta$  and  $\log[N\text{ ii}]/H\alpha$  optical emission line ratios to separate the sources into star forming galaxies, composites, and AGN classes. Additionally, using  $\log[S\text{ ii}]/H\alpha$  and  $\log[O\text{ i}]/H\alpha$  line ratios help to further classify the sources into Seyferts, LINERs, and composites, which are shown in Fig. 12.

For our analysis, we have optical emission line data for 46 262 galaxies from the HECATE-SDSS sample. In Fig. 11, we show the BPT diagram for these galaxies, where the colour coding is scaled by the number of galaxies in each 2D bin (points) and given by the colour bar at right. The curves separating the star-forming, composite, and AGN are marked and labelled.

Using the code provided by Stampoulis et al. (2019) and their definitions, we classified the galaxies using their optical emission line ratios ( $\log[N\text{ ii}]/H\alpha$ ,  $\log[S\text{ ii}]/H\alpha$ ,  $\log[O\text{ i}]/H\alpha$ , and  $\log[O\text{ iii}]/H\beta$ ). We show the histogram of the classifications in Fig. 12: the shaded regions show all the galaxies, notwithstanding of





**Figure 12.** Histogram of galaxy classifications based on the optical emission line ratios, available for 73 percent (46 262 sources) of the HECATE-SDSS sample. For clarity, the inset provides a zoom-in to the y-axis for the shorter bins. The hatched and solid histograms show the input versus the detected sample. Galaxies (blue) make up the majority (81 percent) of the sample, compared to 5.2, 4.1, and 9.8 percent from Seyferts (green), LINERs (yellow), or composites (purple). However, galaxies only comprise 22.6 per cent of the detections. Significant contamination may be produced by Seyferts (23.4 per cent) and LINERs (13.7 per cent), while composite sources appear to be the dominant (40.2 per cent) type to be detected.

being detected, with the percentages for each category marked in the main figure. The solid regions show the  $3\sigma$ -detected galaxies with their composition stated in the inset. Based on the optical emission line diagnostics, we found that the majority of the galaxies in our detected sample (solid histograms) appear to have significant AGN (i.e. Seyfert, LINER) contribution. Therefore, we estimated the total 0.5–10 keV X-ray luminosities, including potential contributions due to AGN, by performing the following Monte Carlo simulation, repeated for each galaxy 500 times:

(i) The Stampoulis et al. (2019) code provides probabilities for the different classifications (i.e. galaxy, Seyfert, LINER, and composite) based on the optical emission line ratios. We used these probabilities as statistical weights ( $f_{\text{gal}}$ ,  $f_{\text{S}}$ ,  $f_{\text{L}}$ ,  $f_{\text{C}}$  for the Seyfert, LINER, and Composite classes, respectively) when summing, as given in the equation below (equation 10).

(ii) We drew 2–10 keV X-ray luminosities randomly from XLFs for each classification type, measured by She, Ho & Feng (2017) for X-ray sources classified as Seyferts, LINERs, and composites. These are designated as  $L_{\text{X,S}}$ ,  $L_{\text{X,L}}$ , and  $L_{\text{X,C}}$ , where the subscripts match the ones above.

(iii) To maintain consistency with the methods described in Section 3, we convert the luminosities into unabsorbed fluxes, using the galaxy’s distance, and multiply this by 50 per cent to account for the expected flux within the PSF.

(iv) We determine conversion factors from 2 to 10 keV to 0.5–10 keV using XSPEC with a photon index of  $\Gamma = 1.8$  for the AGN components, as given in She et al. (2017), and the galaxy absorption used for Section 3. Therefore we determine the 0.5–10 keV model

flux for each component:  $F_{\text{X,S}}$  (Seyfert),  $F_{\text{X,L}}$  (LINER), and  $F_{\text{X,C}}$  (composite).

(v) The 0.5–10 keV X-ray flux is a linear combination of the separate contributions, where the galaxy contribution ( $f_{\text{X,gal}}$ ) was computed as described by equation (3) and in Section 3:

$$F_{\text{X}} = f_{\text{gal}} F_{\text{X,gal}} + f_{\text{S}} F_{\text{X,S}} + f_{\text{L}} F_{\text{X,L}} + f_{\text{C}} F_{\text{X,C}} \quad (10)$$

(vi) We measure the SNR from the XSPEC model counts and background as described in Section 3.2.

We note that the She et al. (2017) XLFs likely suffer from selection bias and incompleteness, and as discussed above, predicting the X-ray emission based on the optical-emission line-defined classes may be problematic. Therefore these estimates are intended to offer a rough comparison between the scenario presented in the previous Section 2.5, where 100 percent of the X-ray emission is assumed to come from XRBs and the hot ISM, and this one, where we add the contribution from accreting supermassive black holes, remaining mindful of the applied assumptions and caveats.

## 4 RESULTS AND DISCUSSION

Based on our analysis of 60 212 galaxies drawn from the HECATE-SDSS and HECATE-IRAS samples, we expect to detect 4 029 galaxies, selected per methods described throughout Section 2, with *eROSITA* based on the analytic simulation of pure galaxies (Sections 3.1–3.3). These estimates use very basic scaling relations for the estimating the X-ray emission based on SFR and stellar mass. On average, this scaling relation applies to the majority of galaxies in our sample and is sufficient for providing estimates for the predicted numbers of galaxies. Recent work has shown that the galaxy-to-galaxy scatter from these relations may be due to physical differences between the galaxies, i.e. stellar ages (Fragos et al. 2008; Lehmer et al. 2017, 2019) or metallicity (Basu-Zych et al. 2013a; Fragos et al. 2013; Douna et al. 2015; Brorby et al. 2016; Lehmer et al. 2016), yet this level of detail is beyond the scope of this paper and does not affect the overall predictions. However, our estimates indicate that the 4-yr *eROSITA* survey will detect sufficient numbers of galaxies to allow detailed investigation of X-ray emission as a function of e.g. stellar age and metallicity.

Noting that the number of predictions depends on our detection (i.e. signal-to-noise) threshold, we present Table 3, which offers a breakdown by galaxy type for three different SNR cuts: 5, 3, and  $2.5\sigma$ . The area subtended by the HECATE-SDSS and HECATE-IRAS samples are 8 500 and 39 189 deg<sup>2</sup>, respectively. By extrapolating our analysis of galaxies selected by the methods detailed in Section 2 to the entire sky by simply scaling the area to the entire sky, we estimate that  $\gtrsim 16\,000$  galaxies may be potentially detected ( $3\sigma$ ) over the entire sky. Columns 3 and 5 provide estimates for the different types and detection thresholds.

Based on the SIXTE simulations (Section 3.4), we expect 3 318 galaxies with  $\text{SNR} \geq 3$ . For comparison, we extrapolate this number over the full sky to provide the lower limit estimate of 14 956 galaxies.

Fig. 13 shows how the galaxies in the *eROSITA* survey expect to compare in terms of distance ( $D$ ) and luminosity (0.5–7 keV;  $L_{\text{X}}$ ) with those from other X-ray surveys. The locus of the analytically simulated *eROSITA* detections is shown in the background along with local star-forming galaxies (blue stars; Mineo et al. 2012a), including luminous infrared galaxies (LIRGs, shown as blue triangles; Lehmer et al. 2010), and early-type galaxies (red squares; Kim & Fabbiano 2013). We show the distant ( $D \gtrsim 400$  Mpc) galaxies from the 7 Ms *Chandra* Deep Field-South (CDFs; Luo et al. 2017) as grey points.

**Table 3.** Predicted numbers for galaxies detected by *eROSITA*. Columns (2) and (4): The number of 0.5–10 keV detected galaxies (and percentage of 60 212 input galaxies) based on the analytic simulation of pure galaxies (Sections 3.1–3.3, assuming no AGN contribution). Columns (3) and (5): The number of galaxies extrapolated over the full sky based on scaling Column (2) or (4) by the area subtended by the samples (8 500 and 39 189 deg<sup>2</sup> for HECATE-SDSS and HECATE-IRAS) to the entire sky (41 252 deg<sup>2</sup>).

(1)	eRASS:1		eRASS:8	
	Simulated # (per cent)	Full sky #	Simulated # (per cent)	Full sky #
(2)	(3)	(4)	(5)	
$5\sigma$				
Total	134 (0.2)	491	1 659 (2.8)	6 873
Early-type	115 (0.2)	467	1 412 (2.3)	6 321
Late-type	19 (0.0)	24	247 (0.4)	553
$3\sigma$				
Total	467 (0.8)	1 924	4 029 (6.7)	16 437
Early-type	406 (0.7)	1 796	3 187 (5.3)	14 414
Late-type	61 (0.1)	129	842 (1.4)	2 023
$2.5\sigma$				
Total	743 (1.2)	3 051	5 269 (8.8)	21 246
Early-type	647 (1.1)	2 847	3 995 (6.6)	18 081
Late-type	96 (0.2)	204	1 274 (2.1)	3 165

We delineate the average *eROSITA* sensitivity limits for the first release of the all-sky survey (eRASS:1; dashed line), the full 4-yr survey (eRASS:8; solid line), and that at the poles after the 4-yr survey (eRASS:8 poles; dotted). In this  $L_X$ – $D$  space, *eROSITA* galaxies occupy a poorly explored region, probing galaxies that are more distant and luminous than most of the nearby populations of galaxies.

As shown in the figure, very few galaxies appear left of the eRASS:1 (dashed) line and those that are detected must be unusually luminous ( $L_X \gtrsim 4 \times 10^{40}$  erg s<sup>−1</sup>) for our sample of 50–200 Mpc galaxies. Nearby (<50 Mpc) galaxies, like many of the marked galaxies, were excluded from our analysis but could be potentially detected. Additionally, the first data release (eRASS:1) could include detections from incredibly rare and unique galaxies with high SFRs, like the LIRGs and low-metallicity, high-redshift analogues (e.g. VV114 or Haro 11, see Basu-Zych et al. 2013a).

However, the majority (85 per cent) of the galaxies in our analysis require deeper exposures. Therefore, owing to its wider area coverage and sensitivity limits, we expect that the *eROSITA* 4-yr survey will be most significant in expanding our capabilities to study the X-ray emission from the complete sample of galaxies, including rare and unique populations.

Most (74 per cent) of these detected galaxies are early types, based on our designation by sSFR ( $<10^{-10.6} M_\odot \text{ yr}^{-1}/M_\odot$ ), where the X-ray emission from diffuse gas dominates the XRB component (see Fig. 7). We note that using sSFR to separate the LMXB- and HMXB-dominated galaxies has been useful for interpreting the X-ray emission from normal galaxies (Mineo et al. 2012a; Lehmer et al. 2016).<sup>15</sup> However, we caution that this classification may yield different results when comparing with other methods of classification

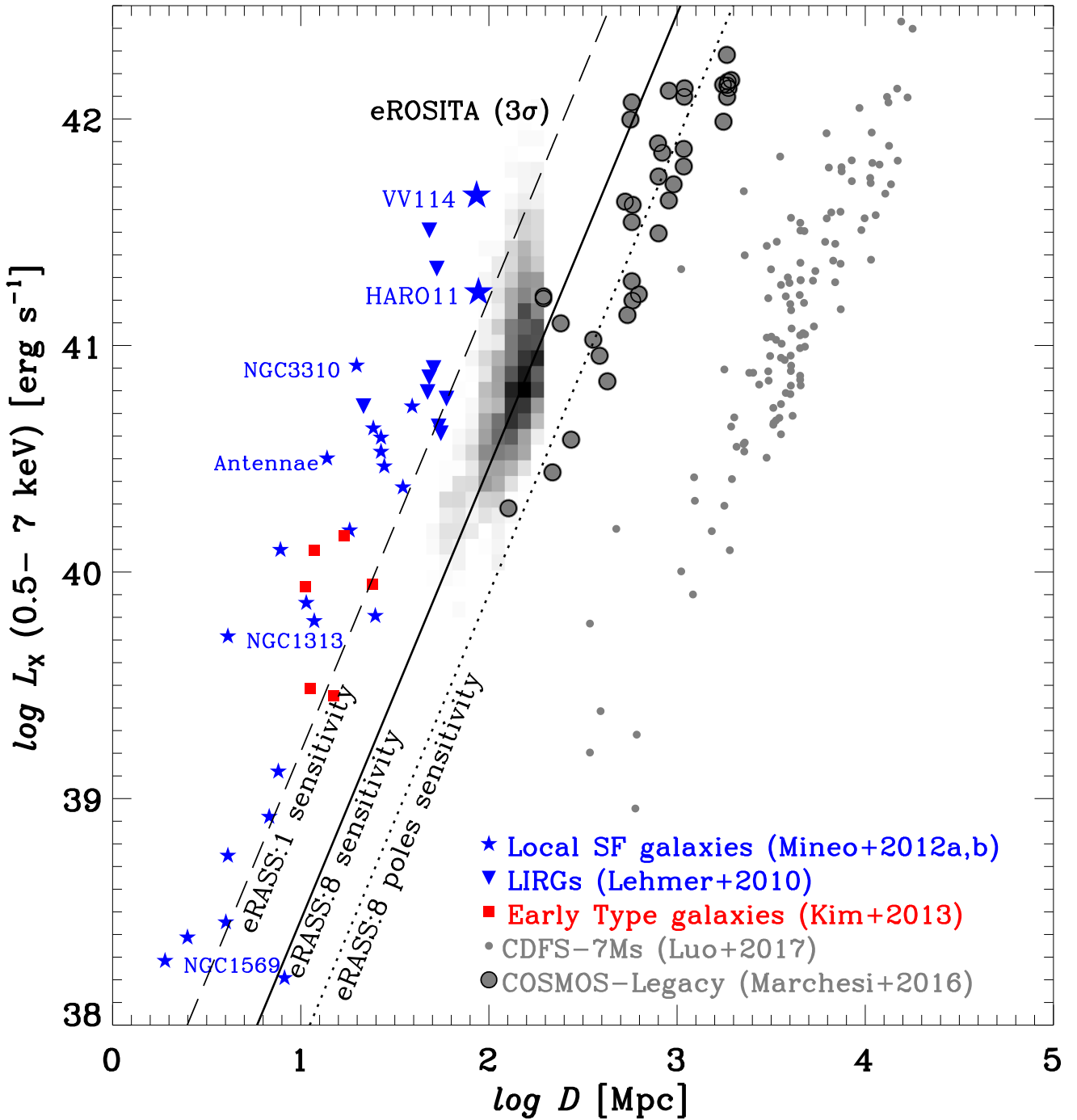
(e.g. morphological or using SED templates). Referring to Fig. 6, we find that detections of late-type galaxies, appearing to the right side of the dashed blue line, come mostly from the IRAS sample (orange), which has a larger fraction of nearby (<100 Mpc), high SFR galaxies than the SDSS sample (magenta), as panels (a) and (b) of Fig. 4 illustrate. In Fig. 14, we show the number counts expected from galaxies, based on our sample selection, divided into early and late types in the middle and right-hand panels. The regions mark the  $1\sigma$  uncertainties based on the Monte Carlo realizations described in Section 3.3. The uncertainties are dominated by the hot gas scaling relations in the early types (middle) and, therefore, total (left-hand panel) galaxies. The prediction by Prokopenko & Gilfanov (2009), scaled to the area of our survey, is shown in the middle and right-hand panels as a grey dashed line for comparison. As discussed above, our galaxy type (i.e. early versus late) classification is quite different from theirs, which is based on spectral templates. Their estimate is within the error region of our estimate and we expect that the main difference is driven by the different classification methods since it appears that our analysis has fewer late types compared to their predictions, but the estimate for the early types shows excellent agreement. As discussed in Section 3.5, some of these galaxies may have significant contributions from AGN activity (i.e. identified as Seyferts, LINERs, and composites). For the sources with measured optical emission lines, we separate  $\log N$ – $\log S$  into types in Fig. 15. The coloured curves mark the contributions from the separate classes: ‘Galaxies’ (solid cyan), Seyferts (long-dashed green), LINERs (thick yellow), and composites (purple). The thick navy line shows the sum of these contributions, which can be compared with the solid black line, which marks the ‘Total sample’ curve from Fig. 14.

At right, each class is shown separately with the same  $x$ - and  $y$ -axes as the main plot, with regions outlining the  $1\sigma$  uncertainties based on the Monte Carlo realizations described in Section 3.5. For comparison, we show the assumption from Fig. 14, where we assumed that all of the emission in these sources arises from galaxy processes alone (i.e. no AGN contribution) as black curves (and grey regions, on the right-hand panels).

Based on this analysis, the composites are dominating at and above the *eROSITA* sensitivity limit and Seyferts appear to be dominating at the brightest end of the flux distribution. We remind readers that the goal here is to provide a rough estimate of the AGN contribution, based on optical-line classifications, and caution against overinterpreting these results, given the caveats discussed in Section 3.5. However, the point here is that the pure galaxy curve (cyan) in Fig. 15 is likely a lower limit due to additional contributions from accreting supermassive black holes in some of the galaxies. Therefore, the measurement of optical emission lines and careful assessment of AGN contamination will be important for *eROSITA* studies of normal galaxies.

We note a few caveats about the predictions of normal galaxies detected by the *eROSITA* survey. Primarily, due to the conservative approach applied in this study, we expect that the numbers presented in this paper are lower limit predictions of *eROSITA* detected galaxies. The selection functions are complex and largely unknown for the parent galaxy samples from which our input sample was drawn, complicating our understanding of the completeness in our study. However, these catalogues contain the largest possible compilation of nearby galaxies and incompleteness is likely only at the faintest end of the galaxy population (e.g. ultra-diffuse, low surface brightness, and/or dwarf galaxies), where the predicted X-ray emission is too faint to yield *eROSITA* detections. Therefore if *eROSITA* ultimately detects significantly more galaxies than predicted, this would alter our understanding substantially of either the underlying galaxy

<sup>15</sup>We note that  $\log(\text{sSFR}) = -10.1 M_\odot \text{ yr}^{-1}/M_\odot$  has been used by these studies to separate LMXB- and HMXB-dominated galaxies. However, this value is fairly extreme and the HMXB-dominated galaxies correspond to very recently star-forming or starburst galaxies, rather than typical late-types. Therefore we use  $-10.6 M_\odot \text{ yr}^{-1}/M_\odot$  to separate the two types, which in Fig. 6 appears to be close to the centre of the bimodal distribution in  $\log(\text{sSFR})$ .

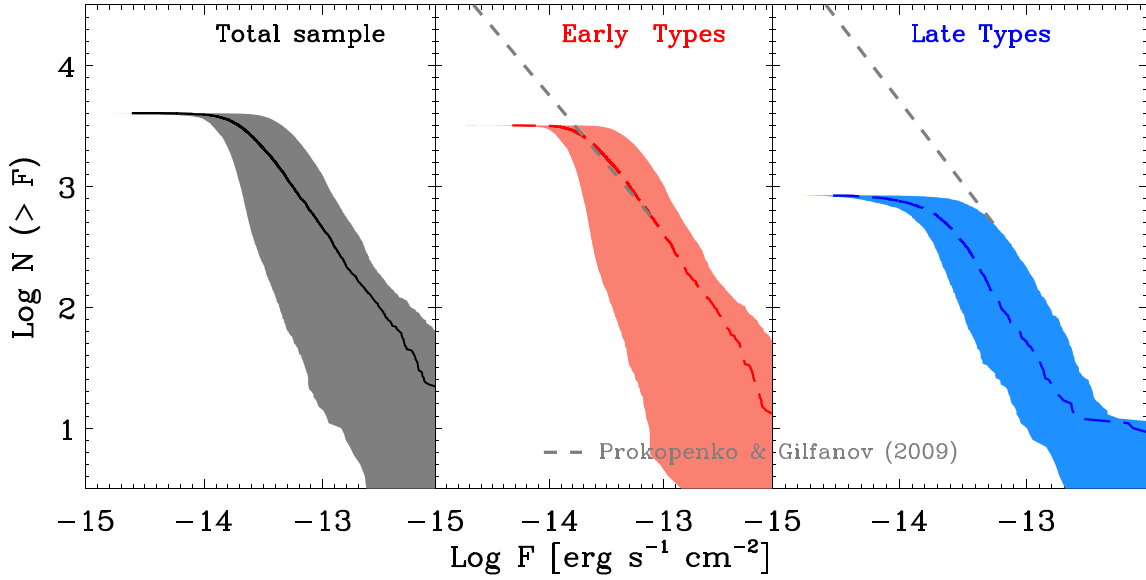


**Figure 13.**  $\log L_X$  versus distance for our *eROSITA* simulated galaxies, shown in the background, compared to nearby star-forming (blue stars; Mineo et al. 2012a) and early-type (red squares; Kim & Fabbiano 2013) galaxies. We mark a few well-known galaxies with their names for reference. Luminous Infrared galaxies (LIRGs) are shown as blue triangles. At further distances, the grey points mark galaxies from the COSMOS Legacy Survey (large encircled symbols; Marchesi et al. 2016) and *Chandra* Deep Field - South 7 Ms survey (small points; Luo et al. 2017). The dotted, dashed, and dash-dotted lines mark the average *eROSITA* sensitivity limits for the single all-sky survey epoch ( $\approx 250$  s), 4-yr all-sky survey ( $\approx 2$  ks), and 4-yr survey at the poles ( $\approx 20$  ks), respectively. Based on this representation, we find that nearly all of the well-studied nearby galaxies will be accessible within the first *eROSITA* data release.

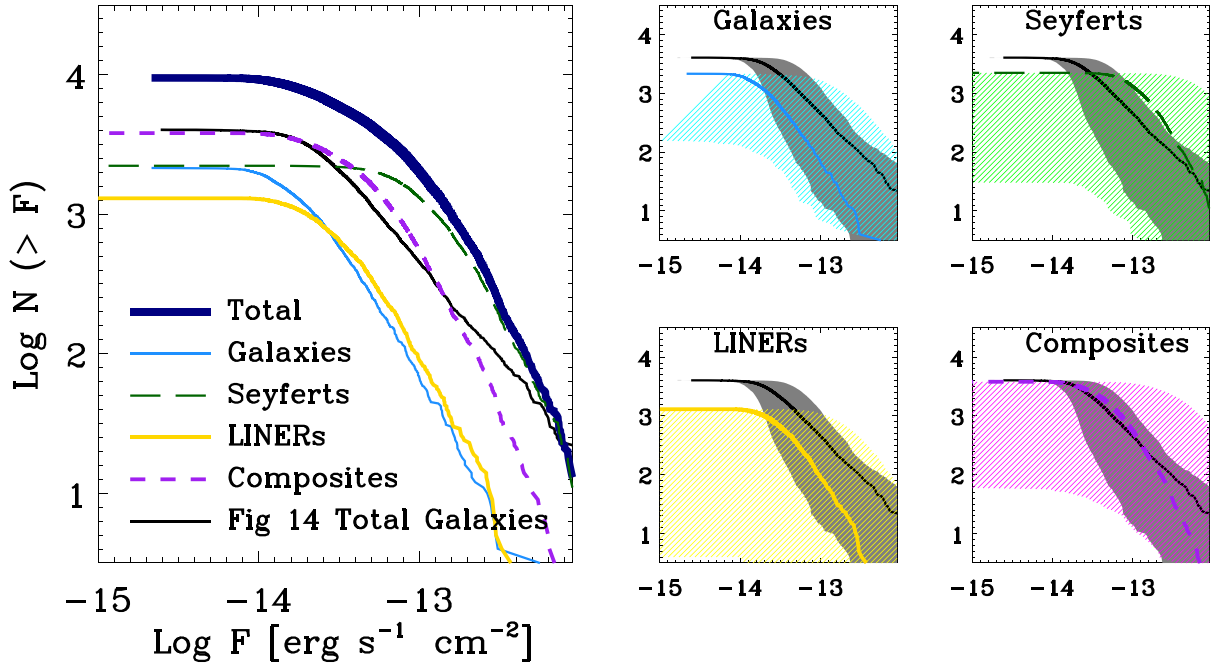
population, as yet undetected at other wavelengths, or of the X-ray emission from faint galaxies.

In the attempt to study a clean and well-understood sample of galaxies, we have restricted the study to unresolved galaxies by

selecting a distance cut  $D > 50$  Mpc. Therefore, individual XRBs in Local Group and confused XRB populations at  $1 \text{ Mpc} \gtrsim D \gtrsim 50 \text{ Mpc}$  will also be detected in the *eROSITA* survey. For example, all the marked galaxies in Fig. 13 would likely be detected and many



**Figure 14.**  $\log N$ – $\log S$  for all galaxies (left-hand panel), early types (middle), and late types (right-hand panel), showing the predicted distribution of fluxes for the  $3\sigma$  detections based on the analytic simulation (Sections 3.1–3.3). The shaded regions mark the expected scatter based on the Monte Carlo simulations described in Section 3.3. The dashed grey line shows the Prokopenko & Gilfanov (2009) predictions for their sample of early and late types. We note that we use the sSFR to separate these two types i.e.  $\log \text{sSFR} > -10.6$  for late-type and  $< -10.6$  for early-type. Galaxies are assumed to be pure galaxies, consisting of XRBs and hot X-ray emitting gas, with no potential contribution from AGN (in contrast to Fig. 15).



**Figure 15.**  $\log N$ – $\log S$  for the  $3\sigma$ -detected sample, based on the analytic simulation (Sections 3.1–3.3), divided into their most probable classification: galaxies (blue), Seyfert (dashed green), LINERs (thick yellow), composites (thick dashed purple). The left-hand panel shows the summed contributions as a thick navy line to compare with the ‘Fig 14 Total Galaxies’ (i.e. including early and late types) curve shown in black from Fig. 14, which assumed X-ray emission from solely hot gas and XRBs. On the right, we divide the classification types, as labelled, into separate figures to show lines and regions corresponding to the mean and  $1\sigma$  distributions based on the Monte Carlo analysis described in Section 3.5. For all panels, the axes have identical ranges to simplify comparison and grey regions and solid black lines provide the scatter and curves from the galaxy-only interpretation shown as ‘Total sample’ in Fig. 14.

others within our lower distance cut, at  $D < 50$  Mpc. Additionally, our sample excluded galaxies within groups or clusters due to the complicated diffuse hot gas contributions. Finally, we have ignored second-order corrections to the X-ray emission scaling relations that may

elevate  $L_X/\text{SFR}$  due to low (sub-solar) metallicities, e.g. in metal-poor starbursts. While predicting the contribution from these galaxies is beyond the scope of this paper, these topics would lend well to the wide-survey and sensitivity capabilities of the 4-yr *eROSITA* survey.



Our study of the normal galaxy population accessible by *eROSITA* highlights several avenues for future galaxy science to which *eROSITA* will contribute, mainly by significantly increasing the numbers of detected galaxies and serving as a pathfinder for rare and bright galaxies. Data from upcoming observatories, such as The Large Synoptic Survey Telescope (LSST) and Dark Energy Survey (DES), can be combined with *eROSITA* to offer the most complete multiwavelength survey of the Universe to date.

## 5 CONCLUSION

The estimates, presented in this paper, are based on the zeroth-order scaling relations. However, a major goal of X-ray studies of normal galaxies is to characterize the nature of the galaxy-to-galaxy scatter observed in these relations. *eROSITA*, by boosting the number of X-ray detected galaxies current by a factor of  $\sim 100$ , will offer the statistical sample required to systematically study the effects of X-ray emission on other galaxy properties such as metallicity and stellar ages.

Based on its expected 4-yr survey coverage and sensitivity, *eROSITA* is expected to significantly increase the number of detected galaxies and further our understanding of the X-ray emission produced in galaxies. In this study, we start from available multiwavelength catalogues, HECATE-SDSS and HECATE-IRAS, to select galaxies between 50 and 200 Mpc. We focus on galaxies at these distances to apply scaling relations that estimate the X-ray emission based on galaxy-wide SFRs and stellar masses, and because at these distances *eROSITA* will likely have the most relevant impact on X-ray galaxy studies. Based on this analysis, we summarize our main results as follows:

(i) Of the 60 212 galaxies in our final sample, selected as described in Section 2, *eROSITA* will detect 3 318 ( $\sim 5$  per cent) galaxies at  $> 3\sigma$  significance. Extrapolating over the full sky, we predict that *eROSITA* will detect 14 956 galaxies, which will be divided equally between the *eROSITA*-DE and *eROSITA*-RU data.

(ii) Most ( $\sim 70$  per cent) of these detections will be galaxies with  $\log \text{sSFR} < -10.6$ , which we designate as early-type galaxies for our study.

(iii) Due to our conservative approach, our estimate is a lower limit for the following reasons:

(a) *eROSITA* is also likely to detect nearby ( $< 50$  Mpc) galaxies or their spatially resolved components (i.e. XRBs).

(b) We have excluded galaxies residing in or near groups or clusters from our analysis due to the complicated hot gas contribution from these environments.

(c) We did not apply the metallicity-dependent scaling of X-ray emission with SFR, which could lead to higher X-ray luminosities and, therefore, lead to potentially more detections of low-metallicity (and late-type) galaxies.

(iv) As shown in Fig. 13, the *eROSITA* 4-yr survey probes a unique region in terms of distance and X-ray luminosity compared to other galaxy studies. While the eRASS:1 data (i.e. first data release including 6 months of survey data) may discover extremely rare, X-ray luminous galaxies, the full 4-yr survey will ultimately offer the most complete and statistically robust X-ray study of galaxies.

(v) Given the expected number counts (see Fig. 15), multiwavelength follow-up observations, especially optical spectroscopy, will be required to help distinguish normal galaxies from other classes of galaxies (e.g. Seyferts, LINERs, and composites).

## ACKNOWLEDGEMENTS

We thank the referee for helpful suggestions that improved the clarity of this manuscript. We gratefully acknowledge Violet Replicon for her contributions to this work, Mackenzie Jones for sharing code and data on AGN fractions, Thomas Reiprich for providing useful suggestions that significantly improved the analysis, and Philipp Weber for useful discussions and provision of some of the SIMPUT catalogues provided here. We acknowledge funding support from NASA GSFC through its Internal Research And Development (IRAD) program. This work was partially funded by the Bundesministerium für Wirtschaft und Technologie based on a resolution of the German Parliament through Deutsches Zentrum für Luft- und Raumfahrt grant 50 QR 1603. We acknowledge use of public data from the Sloan Digital Sky Survey. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III website is <http://www.sdss3.org/>. This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

## DATA AVAILABILITY

Most of the data underlying this article is available via publicly available catalogues, e.g. HyperLEDA (<http://leda.univ-lyon1.fr/leda/param/objtype.html>), the GALEX–SDSS–WISE Legacy Catalogue (<https://salims.pages.iu.edu/gswlcl/>; see Salim et al. 2016) and the Revised IRAS-FSC Redshift Catalogue (<http://www.astro.dur.ac.uk/~lwang83/RIFSCz.tar.gz>; see Wang et al. 2014). Any additional data request based on this article can be requested by contacting the corresponding author.

## REFERENCES

- Agostino C. J., Salim S., 2019, *ApJ*, 876, 12
- Ahn C. P. et al., 2014, *ApJS*, 211, 17
- Alam S. et al., 2015, *ApJS*, 219, 12
- Anders E., Grevesse N., 1989, *Geochim. Cosmochim. Acta*, 53, 197
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
- Basu-Zych A. R. et al., 2013a, *ApJ*, 762, 45
- Basu-Zych A. R. et al., 2013b, *ApJ*, 774, 152
- Bauer F. E., Alexander D. M., Brandt W. N., Schneider D. P., Treister E., Hornschemeier A. E., Garmire G. P., 2004, *AJ*, 128, 2048
- Bell E. F. et al., 2005, *ApJ*, 625, 23
- Bianchi L., Herald J., Efremova B., Girardi L., Zabot A., Marigo P., Conti A., Shiao B., 2011, *Ap&SS*, 335, 161
- Boller T., Meurs E. J. A., Brinkmann W., Fink H., Zimmermann U., Adorf H.-M., 1992, *A&A*, 261, 57
- Brorby M., Kaaret P., Prestwich A., Mirabel I. F., 2016, *MNRAS*, 457, 4081
- Cappelluti N. et al., 2007, *ApJS*, 172, 341
- Chary R., Elbaz D., 2001, *ApJ*, 556, 562
- Cutri R. M. et al., 2003, *VizieR Online Data Catalog*. p. II/246

- Cutri R. M. et al., 2014, *VizieR Online Data Catalog*, p. II/328
- Das A., Mesinger A., Pallottini A., Ferrara A., Wise J. H., 2017, *MNRAS*, 469, 1166
- Dauser T. et al., 2019, *A&A*, 630, 66
- David L. P., Jones C., Forman W., 1992, *ApJ*, 388, 82
- Douna V. M., Pellizza L. J., Mirabel I. F., Pedrosa S. E., 2015, *A&A*, 579, 44
- Dwelly T. et al., 2017, *MNRAS*, 469, 1065
- Eracleous M., Shields J. C., Chartas G., Moran E. C., 2002, *ApJ*, 565, 108
- Fabbiano G., 1989, *ARA&A*, 27, 87
- Fabbiano G., Kim D.-W., Trinchieri G., 1992, *ApJS*, 80, 531
- Finoguenov A. et al., 2015, *A&A*, 576, A130
- Flohic H. M. L. G., Eracleous M., Chartas G., Shields J. C., Moran E. C., 2006, *ApJ*, 647, 140
- Fragos T. et al., 2008, *ApJ*, 683, 346
- Fragos T. et al., 2013, *ApJ*, 764, 41
- Georgakakis A. E., Georgantopoulos I., Basilakos S., Plionis M., Kolokotronis V., 2004, *MNRAS*, 354, 123
- Georgakakis A. E., Chavushyan V., Plionis M., Georgantopoulos I., Koulouridis E., Leonidaki I., Mercado A., 2006, *MNRAS*, 367, 1017
- Georgantopoulos I., Georgakakis A., Koulouridis E., 2005, *MNRAS*, 360, 782
- Greig B., Mesinger A., 2017, *MNRAS*, 472, 2651
- Hornschemeier A. E. et al., 2003, *AJ*, 126, 575
- Huchra J. P. et al., 2012, *ApJS*, 199, 26
- Jarrett T. H. et al., 2011, *ApJ*, 735, 112
- Jones M. L., Hickox R. C., Black C. S., Hainline K. N., DiPompeo M. A., Goulding A. D., 2016, *ApJ*, 826, 12
- Kim D.-W., Fabbiano G., 2013, *ApJ*, 776, 116
- Kim D.-W., Fabbiano G., 2015, *ApJ*, 812, 127
- Kim D.-W. et al., 2006, *ApJ*, 644, 829
- Kroupa P., 2001, *MNRAS*, 322, 231
- LaMassa S. M., Georgakakis A., Vivek M., Salvato M., Tasnim Ananna T., Urry C. M., MacLeod C., Ross N., 2019, *ApJ*, 876, 50
- Lang D., Hogg D. W., Schlegel D. J., 2016, *AJ*, 151, 36
- Lehmer B. D. et al., 2012, *ApJ*, 752, 46
- Lehmer B. D. et al., 2014, *ApJ*, 789, 52
- Lehmer B. D. et al., 2016, *ApJ*, 825, 7
- Lehmer B. D. et al., 2017, *ApJ*, 851, 11
- Lehmer B. D. et al., 2019, *ApJS*, 243, 3
- Lehmer B. D., Alexander D. M., Bauer F. E., Brandt W. N., Goulding A. D., Jenkins L. P., Ptak A., Roberts T. P., 2010, *ApJ*, 724, 559
- Li T. P., Ma Y. Q., 1983, *ApJ*, 272, 317
- Luo B. et al., 2017, *ApJS*, 228, 2
- Maccarone T. J., Lehmer B. D., Leyder J. C., Antoniou V., Hornschemeier A., Ptak A., Wik D., Zezas A., 2014, *MNRAS*, 439, 3064
- Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, *A&A*, 570, A13
- Marchesi S. et al., 2016, *ApJ*, 830, 100
- Merloni A. et al., 2012, preprint (arXiv:1209.3114)
- Mesinger A., Ferrara A., Spiegel D. S., 2013, *MNRAS*, 431, 621
- Mesinger A., Ewall-Wice A., Hewitt J., 2014, *MNRAS*, 439, 3262
- Mineo S., Gilfanov M., Sunyaev R., 2012a, *MNRAS*, 419, 2095
- Mineo S., Gilfanov M., Sunyaev R., 2012b, *MNRAS*, 426, 1870
- Molina M., Eracleous M., Barth A. J., Maoz D., Runnoe J. C., Ho L. C., Shields J. C., Walsh J. L., 2018, *ApJ*, 864, 90
- Norman C. et al., 2004, *ApJ*, 607, 721
- Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, *MNRAS*, 443, 678
- Planck Collaboration et al., 2014, *A&A*, 571, A1
- Predehl P. et al., 2016, in *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*, Proceedings of the SPIE, Vol. 9905. p. 99051K
- Prokopenko I. G., Gilfanov M. R., 2009, *Astron. Lett.*, 35, 294
- Ranalli P., Comastri A., Setti G., 2003, *A&A*, 399, 39
- Salim S. et al., 2016, *ApJS*, 227, 2
- Seibert M. et al., 2012, in *American Astronomical Society Meeting Abstracts* #219. p. 340.01
- She R., Ho L. C., Feng H., 2017, *ApJ*, 835, 223
- Snowden S. L. et al., 1997, *ApJ*, 485, 125
- Stampoulis V., van Dyk D. A., Kashyap V. L., Zezas A., 2019, *MNRAS*, 485, 1085
- Strauss M. A. et al., 2002, *AJ*, 124, 1810
- Tajer M., Trinchieri G., Wolter A., Campana S., Moretti A., Tagliaferri G., 2004, *Mem SA It Suppl.*, 5, 287
- Taylor M., 2011, *Astrophysics Source Code Library*, record ascl:1101.010
- Türler M., Chernyakova M., Courvoisier T. J.-L., Lubiński P., Neronov A., Produit N., Walter R., 2010, *A&A*, 512, A49
- Voges W. et al., 1999, *A&A*, 349, 389
- Wang L., Rowan-Robinson M., Norberg P., Heinis S., Han J., 2014, *MNRAS*, 442, 2739
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- Zibetti S., Charlot S., Rix H.-W., 2009, *MNRAS*, 400, 1181
- Zimmermann H.-U., Boller T., Döbereiner S., Pietsch W., 2001, *A&A*, 378, 30

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