UNIVERSITY OF CRETE

DEPARTMENT OF PHYSICS

PhD THESIS

Populations of ultraluminous X-ray sources in nearby galaxies

Konstantinos Kovlakas

Advisor: Professor Andreas Zezas

Heraklion, 2020

University of Crete
Department of Physics

PhD THESIS

Doctor of Philosophy
in
ASTROPHYSICS

Defended by
Konstantinos Kovlakas

Populations of ultraluminous X-ray sources in nearby galaxies

Committee

Professor	Andreas Zezas	University of Crete
Professor Emeritus	Nikolaos Kylafis	University of Crete
Professor	Vasileios Charmandaris	University of Crete
Professor	Iosif Papadakis	University of Crete
Associate Professor	Konstantinos Tassis	University of Crete
Assistant Professor	Anastatsios Fragos	University of Geneva
Senior Researcher	Pablo Reig	Institute of Astrophysics / FORTH

Date of the defence:
15/12/2020

This page intentionally left blank

Abstract

Ultraluminous X-ray sources (ULXs) are the most energetic non-explosive stellar systems, with luminosities exceeding the Eddington limit for stellar-mass black hole, $\sim 10^{39}$ erg s^{-1}. They are formed by a compact object (black hole or a neutron star), and a companion star from which material is extracted, forming an accretion disk, and eventually falling onto the compact object. The release of the gravitational energy of the accreted material powers the extreme luminosities of ULXs, which sometimes dominate the X-ray output of their host galaxies.

Since ULXs are rare – typically one per galaxy – they are predominantly found at large distances. Therefore, only a handful of systems have been studied thoroughly, with the general properties of ULXs being still a matter of debate. The nature of the accretor or the donor star, and the exact structure of the accretion disk are not fully understood. Furthermore, their puzzling super-Eddington luminosities challenge our understanding of the physics at extreme accretion rates, and has implications for the effect of ULXs in the evolution of galaxies, and the intergalactic medium in the early Universe.

The study of ULXs in the context of the host galaxies can constrain models for their formation and evolution, and consequently provide input for the progenitors of endpoints of massive binary systems, such as short gamma-ray bursts, and gravitational-wave sources. In this respect, demographic studies of ULXs probe the rate of ULXs in galaxies, and its connection to their stellar population parameters, such as star-formation rate, stellar mass, and metallicity.

In this thesis, we create a census of ULX populations in the local Universe, and we revisit their connection with the parameters of the host galaxies.

First, we create a all-sky, value-added catalogue of galaxies at distances up to 200 Mpc, the *Heraklion Extragalactic Catalogue* (*HECATE*), providing all the necessary

information for the detailed study of ULXs in nearby galaxies. Using information from astrophysical databases, and multi-wavelength data, we determine the distances, and the stellar population parameters of the galaxies. Furthermore, we discuss possible applications of the catalogue for other fields of astrophysics, such as the characterisation of sources in multi-wavelength surveys, the identification of possible host galaxies of gravitational-wave sources, and gamma-ray bursts, etc.

By associating the *HECATE* with the *Chandra Source Catalog 2.0*, we deliver a census of ULX populations in galaxies up to a distance of 40 Mpc (at which source confusion is not severe). In 309 galaxies, we find 629 ULX candidates. By statistically correcting for foreground and background interlopers (~20% of the total number of ULX candidates), we estimate the number of ULXs in the sample galaxies, and probe its correlation with star-formation rate, stellar mass, and metallicity. Hence, we provide the tightest constrain on the average scaling of ULXs with star-formation rate (SFR) in late-type galaxies, 0.51 ± 0.06 ULXs per $\mathrm{M_\odot\ yr^{-1}}$. This is close to the expected value from recent high-mass X-ray binary luminosity functions, in line with the expectation that ULXs in spiral galaxies are luminous high-mass X-ray binaries. By accounting for sample differences, we explain the tension with older estimates of the ULX-SFR scaling.

For the first time, we account for the contribution of low-mass X-ray binaries (which scale with stellar mass, $M_\star$), in the ULX population of spiral galaxies, finding $0.45^{+0.06}_{-0.09}\times\frac{SFR}{\mathrm{M_\odot\ yr^{-1}}} + 3.3^{+3.8}_{-3.2}\times\frac{M_\star}{\mathrm{M_\odot}}$. The above scaling depends on the morphological type of the host galaxies, with early spiral galaxies (i.e., S0/a-Sbc) containing a small but non-negligible population of luminous, low-mass X-ray binaries. In addition, we confirm the known anti-correlation of the ULX rate with the metallicity of the host galaxy, and attribute to this effect, the excess of ULXs in low-mass and late-type galaxies in our sample.

Regarding the specific ULX frequency in early-type galaxies, we find 6.3 ± 1.0 ULXs per $10^{12}\mathrm{M_\odot}$, in agreement with previous demographic results, and low-mass X-ray binary luminosity functions, suggesting that ULXs in these galaxies are luminous low-mass X-ray binaries. However, we discover an excess in low–stellar-mass early-type galaxies, which we reproduce using binary population synthesis models, and realistic star-formation histories. Finally, we hint at an excess of ULXs at large galactrocentric distances in elliptical galaxies, indicating possible contribution from globular cluster ULXs, in line with previous studies.

Ultimately, using X-ray images from *Chandra* observations, we simulate the effect of source blending when targeting galaxies at distances from 10 Mpc up to 100 Mpc. We identify the biases of source blending in terms of the slopes of luminosity functions and the number of ULXs. We find that at larger distances ($\gtrsim$60 Mpc), source confusion leads to a systematic overestimation in the number of ULXs. Finally, we highlight the importance of the clumpy nature of star-forming regions, and diffusion emission in the observed number of ULXs, even at lower distances.

This page intentionally left blank

Περίληψη

Οι υπέρλαμπρες πηγές ακτίνων Χ (ΥΠΧ) είναι οι τα πιο ενεργητικά, μη-εκρηκτικά αστρικά συστήματα, με λαμπρότητες που ξεπερνούν το όριο Eddington για μελανές οπές αστρικής μάζας, $\sim 10^{39}$ erg s^{-1}. Σχηματίζονται από ένα συμπαγή αστέρα (μελανή οπή ή αστέρα νετρονίων) και ένα συνοδό αστέρα από τον οποίο μεταφέρεται υλικό, σχηματίζοντας ένα δίσκο προσαύξησης και καταλήγοντας εν τέλει στο συμπαγή αστέρα. Η απελευθέρωση βαρυτικής ενέργειας από την προσαυξανόμενη ύλη τροφοδοτεί τις ακραίες λαμπρότητες των ΥΠΧ, που μερικές φορές κυριαρχούν στη συνολική εκπομπή ακτίνων Χ των γαλαξιών που τις φιλοξενούν.

Καθώς οι ΥΠΧ είναι σπάνιες – τυπικά μία ανά γαλαξία – εντοπίζονται κυρίως σε μεγάλες αποστάσεις. Έτσι, μόνο μερικά από αυτά τα συστήματα έχουν μελετηθεί διεξοδικά, με τις γενικές ιδιότητες του πληθυσμού τους να αποτελούν αμφιλεγόμενο ζήτημα. Η φύση του συμπαγούς αστέρα και του συνοδού του, και η ακριβής δομή του δίσκου προσαύξησης δεν είναι πλήρως κατανοητές. Επιπλέον, οι αινιγματικές τους λαμπρότητες πέραν του ορίου Eddington αποτελούν πρόκληση για τη Φυσική της ταχείας προσαύξησης μάζας, και έχουν συνέπειες για την επίδραση των ΥΠΧ στην εξέλιξη των γαλαξιών και τη διαγαλαξιακή ύλη στα πρώιμα στάδια του Σύμπαντος.

Η μελέτη των ΥΠΧ στα πλαίσια των φιλοξενούντων γαλαξιών τους μπορεί να επιβάλλει περιορισμούς στις παραμέτρους που επηρεάζουν το σχηματισμό και την εξέλιξη τους, και συνεπώς μπορεί να δώσει πληροφορίες για τα τελικά στάδια διπλών αστέρων μεγάλης μάζας, όπως αυτά που δημιουργούν εκλάμψεις ακτίνων-γ βραχείας διάρκειας, και πηγές βαρυτικών κυμάτων. Έτσι, δημογραφικές μελέτες των ΥΠΧ εξετάζουν τη συχνότητα των ΥΠΧ σε γαλαξίες, και

τη σχέση της με τις αστρικές παραμέτρους τους, όπως ο ρυθμός αστρογέννεσης, η αστρική μάζα και η μεταλλικότητα.

Σε αυτήν τη διατριβή, πραγματοποιούμε μία απογραφή των πληθυσμών ΥΠΧ στο τοπικό Σύμπαν, και επανεξετάζουμε τη σύνδεσή τους με τις παραμέτρους των γαλαξιών στους οποίους ανήκουν.

Πρώτα, κατασκευάζουμε έναν κατάλογο γαλαξιών σε αποστάσεις <200 Mpc, τον *Heraklion Extragalactic Catalogue* (*HECATE* ή *Εκάτη*), που παρέχει όλες τις απαραίτητες πληροφορίες για μία εμπεριστατομένη μελέτη των ΥΠΧ σε κοντινούς γαλαξίες. Χρησιμοποιώντας δεδομένα από αστρονομικές βάσεις δεδομένων, και παρατηρήσεις σε διάφορα μήκη κύματος, προσδιορίζουμε τις αποστάσεις και τις αστρικές παραμέτρους των γαλαξιών. Επιπλέον, συζητούμε πιθανές άλλες χρήσεις του καταλόγου για άλλους τομείς της Αστροφυσικής, όπως ο χαρακτηρισμός πηγών από συστηματικές έρευνες σε διάφορα μήκη κύματος, ο εντοπισμός των γαλαξιών στους οποίους βρίσκονται πηγές βαρυτικών κυμάτων, κ.α.

Συνδυάζοντας την *Εκάτη* με τον δεύτερο κατάλογο του *Chandra*, κατασκευάζουμε έναν κατάλογο πληθυσμών ΥΠΧ σε γαλαξίες με αποστάσεις μέχρι 40 Mpc. Βρίσκουμε 629 ΥΠΧ σε 309 γαλαξίες. Διορθώνοντας στατιστικά τη συνεισφορά (~20%) από αστέρια και πηγές υποβάθρου, εκτιμούμε το πλήθος των ΥΠΧ σε κάθε γαλαξία στο δείγμα μας, και μελετούμε τη σχέση του με το ρυθμό αστρογένεσης, την αστρική μάζα και τη μεταλλικότητα. Ως αποτέλεσμα, δίνουμε την πιο ακριβή μέτρηση της αναλογίας των ΥΠΧ σε γαλαξίες μεταγενέστερου τύπου με τον αριθμό αστρογένεσης (SFR), 0.51±0.06 ΥΠΧ ανά $M_\odot\,yr^{-1}$. Αυτή η τιμή είναι πολύ κοντά στα αποτελέσματα πρόσφατα δημοσιευμένων κατανομών λαμπρότητας διπλών αστέρων ακτίνων Χ μεγάλης μάζας. Λαμβάνοντας υπόψη τις διαφορές των δειγμάτων, εξηγούμε τη διαφορά με παλαιότερες μετρήσεις της αναλογίας σε μελέτες που εστίασαν σε ΥΠΧ.

Λαμβάνουμε υπόψη, για πρώτη φορά, τη συνεισφορά των διπλών αστέρων ακτίνων Χ μικρής μάζας (που είναι ανάλογη της αστρικής μάζας, $M_\star$), βρίσκοντας $0.45^{+0.06}_{-0.09} \times \frac{SFR}{M_\odot\,yr^{-1}} + 3.3^{+3.8}_{-3.2} \times \frac{M_\star}{M_\odot}$. Οι παραπάνω συντελεστές εξαρτώνται από τον μορφολογικό τύπο των γαλαξιών, με τους πρώιμους σπειροειδείς (π.χ., S/a-Sbc) να φιλοξενούν ένα μικρό αλλά μη-αμελητέο πληθυσμό από λαμπρούς διπλούς αστέρες ακτίνων Χ μικρής μάζας. Ακόμη, επιβεβαιώνουμε της γνωστή

αρνητική συσχέτιση της συχνότητας των ΥΠΧ με τη μεταλλικότητα, και αποδίδουμε σε αυτή, το πλεόνασμα ΥΠΧ σε γαλαξίες μικρής μάζας ή/και ύστερου μορφολογικού τύπου.

Σχετικά με την ειδική συχνότητα των ΥΠΧ στους γαλαξίες πρώιμου τύπου, βρίσκουμε 6.3 ± 1.0 ΥΠΧ ανά $10^{12}\,M_{\odot}$, σε συμφωνία με προηγούμενα αποτελέσματα, και κατανομές λαμπρότητας διπλών αστέρων ακτίνων Χ μικρής μάζας, υποδεικνύοντας τους τελευταίους ως την προέλευση των πληθυσμών ΥΠΧ σε ελλειπτικούς γαλαξίες. Ωστόσο, ανακαλύπτουμε ένα πλεόνασμα ΥΠΧ σε γαλαξίες πρώιμου τύπου και μικρής μάζας, το οποίο αναπαράγουμε με το συνδυασμό αποτελεσμάτων από προσομοιώσεις διπλών αστέρων, και ιστορίες αστρογένεσης από πραγματικούς γαλαξίες. Επιπλέον, βρίσκουμε ενδείξεις για ένα πληθυσμό ΥΠΧ σε μεγάλες αποστάσεις από τα κέντρα ελλειπτικών γαλαξιών, υποδεικνύοντας πιθανή συνεισφορά από πηγές σε σφαιρωτά σμήνη, όπως έχει αναφερθεί σε προηγούμενες έρευνες.

Τέλος, χρησιμοποιώντας εικόνες από παρατηρήσεις ακτίνων Χ με το *Chandra*, προσομοιώνουμε την επίδραση του φαινομένου της σύγχυσης πηγών, σε αποστάσεις από 10 ως 100 Mpc. Προσδιορίζουμε τα συστηματικά σφάλματα στις κατανομές λαμπρότητας, και στο παρατηρούμενο πλήθος των ΥΠΧ. Βρίσκουμε ότι σε αποστάσεις μεγαλύτερες των 60 Mpc, η σύγχυση των πηγών προκαλεί συστηματική υπερεκτίμηση των ΥΠΧ. Τέλος, τονίζουμε τη σημασία της ‘‘σβολώδους’’ εμφάνισης των περιοχών αστρογένεσης και διάχυτης ακτινοβολίας, στο παρατηρούμενο πλήθος τον ΥΠΧ.

This page intentionally left blank

To Christina Noutsou and Prof. John H. Seiradakis,
who believed in me when I didn't.

This page intentionally left blank

Preface and Acknowledgments

Black holes, neutron stars, and white dwarves were often described as 'dead stars' by my Astronomy professors. The idea that stars have a lifeline was impressive, but also... depressing.

Later, I was amazed to find out that massive stars live short albeit intense lives, and mark their end with spectacular explosions. However, I was inspired by the notion hat the outcome of their swan songs was the enrichment of the space with the ingredients for the next generations of stars, the planets around them, and organic matter. These life-giving lyrics contained the words which gave the Universe a chance to understand itself, as Carl Sagan used to say.

During this dissertation, I learned that massive stars are often found in binaries, and their lives may still continue after their final stage of their evolution. Sometimes, the 'companions' having spent almost all their lives together, they may survive as a system, and interact by exchanging material. The result is spectacular: the companion star revives the, otherwise, 'dead' star, making it shine bright anew, echoing from the most distant corners the Universe.

I want to express my gratitude to Andreas Zezas, who gave me the chance to be inspired once again. His commitment to Science, his passion for Education, and his genuine interest on the well-being of his students, make him the most supportive mentor. I thank him for the sleepless nights helping me, and the numerous times I could count on him, on a personal and a professional level.

I would like to thank the other members of the advisory committee, K. Tassis, and P. Reig, as well as faculty members V. Charmandaris, T. Diaz Santos, N. Kylafis, I. Papadakis, V. Pavlidou, for providing scientific and professional support, and encouraging my career. They have created a stimulating, productive and vibrant environment, which impressed me from my first day in the University of Crete. Their heritage is responsible for the excellence of their former students, and some footsteps heard in the corridors of famous institutes around the world. We are all grateful.

The fulfilment of this dissertation, and my delving in X-ray astronomy wouldn't be possible without the help of K. Anastasopoulou, J. Andrews, A. Basu-Zych, F. Civano, R. D'Abrusco, T. Dauser, K. Doore, R. Eufrasio, G. Fabbiano, T. Fragos, K. Garofali, F. Haberl, A. Hornschemeier, W. Jörn, K. Kouroumpatzakis, B. Lehmer, S.-S., Luidy, A. Ptak, A. Rots, P. Sell, N. Vulic, and B. Williams. I hope for a continuing collaboration with such excellent scientists.

Working into diverse areas of research was particularly stimulating, something that I owe to S. Bavera, C. Berry, P. Marchant, V. Kalogera, S. Coughlin, A. Dotter, D. Misra, J. G. Serra-Perez, Y. Qin, K. Rocha, J. Román-Garza, N. H. Tran, E. Zapartas (binary evolution), G. Korkidis, E. Ntormousi, and T. Tomaras (cosmology), P. Bonfini, S. Willner, M. Ashby, A. Maragkoudakis, T. Jarrett (star-formation).

The quality of the research leading to this thesis was greatly improved by feedback from M. Colom i Bernadich, K. Dage, R. Ebert, D. Lang, D. Makarov, A. Schwope, and V. Replicon, as well as the anonymous referees of the articles.

Special thanks to V. Antoniou, D. Blinov, C. Casadio, F. Del Sordo, J. Kapetanakis, S. Kiehlmann, I. Komis, M. Kopsacheili, E. Kyritsis, I. Liodakis, I. Leonidaki, N. Mandarakas, G.-M., Maragkakis, G. Maravelias, G. Panopoulou, V. Pelgrims, C. Politikakis, R. Skalidis, A. Tritsis, and S. Williams, for sharing experiences and personal moments, as well as T. Kougentakis, G. Paterakis, G. Kapetanakis, A. Steiakaki, and E. Tsentelierou, for making our life easier.

I will never forget my former professors who helped, encouraged and believed in me, M. Plionis, N. Stergioulas, K. Tsiganis, C. Tsagas, L. Vlahos, G. Voyatzis, C. Polatoglou, and N. Fragkis. I am grateful to J. H. Seiradakis† — he will always be remembered as the 'teacher' inspiring my stance towards Astronomy and Science Outreach.

I cannot thank enough my parents, Kostas and Maria, my brother Vasilis, for the continuing support in fulfilling my dreams.

I am fortunate for having Christina Noutsou in my life, sharing these dreams, and inspiring many more despite all the difficulties faced and sacrifices made. No words can express my gratitude.

The scientific results reported in this article are based to a significant degree on data obtained from the Chandra Data Archive and the Chandra Source Catalog provided by the Chandra X-ray Center (CXC). We acknowledge the usage of the HyperLeda database, `TOPCAT` *astronomical software, as well as the cross-match service and Vizier catalgue access tool provided by CDS, Strasbourg, France. We acknowledge 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. 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. 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 web site is http://www.sdss3.org.*

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 617001, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie RISE action, Grant Agreement n. 691164 (ASTROSTAT), the NASA Astrophysics Data Analysis Program 80NSSC20K0444 (B. Lehmer), and CIERA and Northwestern University through a Postdoctoral Fellowship (J. Andrews).

Konstantinos Kovlakas
Heraklion, December 2020

This page intentionally left blank

Declaration

I hereby declare that the work in this thesis is that of the candidate alone, except where indicated below.

Chapter 2: published in the Monthly Notices of the Royal Astronomical Society as: *Kovlakas K., Zezas A., Andrews J. J., Basu-Zych A., Fragos T., Hornschemeier A., and Lehmer B., et al., 2020, MNRAS, 498, 4790.*

Chapter 3: submitted to the Monthly Notices of the Royal Astronomical Society. This paper is co-authored by Zezas A., Andrews J. J., Basu-Zych A., Fragos T., Hornschemeier A., Kouroumpatzakis K., and Lehmer B.

Chapter 4: paper in preparation, co-authored by Zezas A.

This page intentionally left blank

Contents

List of Tables

This page intentionally left blank

List of Figures

Abbreviations

Term	Explanation
AGN	Active galactic nucleus (or nuclei)
BH	Black hole
BNS	Binary neutron star
CI	Confidence interval
CR	Confidence region
CSC	Chandra Source Catalog
EM	Electromagnetic
ETG	Early-type galaxy
f/b	Foreground/background
FIR	Far infrared
GC	Globular cluster
GRB	Gamma-ray burst
GW	Gravitational wave
HECATE	Heraklion Extragalactic Catalogue
HLX	Hyperluminous X-ray source
HMXB	High-mass X-ray binary
IMBH	Intermediate-mass black hole
IGM	Intergalactic medium
IR	Infrared
LF	Luminosity function
LG	Local Group

continued on next page ...

... continued from previous page

Term	Explanation
LMC	Large Magellanic Cloud
LMXB	Low-mass X-ray binary
LTG	Late-type galaxy
MLE	Maximum likelihood estimator
NS	Neutron star
PSF‘	Point spread function
RLO	Roche-lobe overflow
SED	Spectral energy distribution
SFH	Star-formation history
SFR	Star-formation rate
SMBH	Supermassive black hole
SMC	Small Magellanic Cloud
sSFR	Specific star-formation rate
TIR	Total infrared
ULX	Ultraluminous X-ray source
UV	Ultraviolet
VC	Virgo Cluster
XLF	X-ray luminosity function
XRB	X-ray binary

1 Introduction

1.1 X-ray astronomy and accreting binary stars

After the end of World War II, technology once used in combat and intelligence, passed to the sciences, including astronomy. Just like radars were put to use by radio-astronomers, V2 rockets were carrying the first detectors for the study of the X-rays coming from the Sun — the only X-ray source known until 1962. However, the decade of 1960s was devoted to the Moon, and the upcoming manned missions. In the process of studying scattered solar X-ray radiation by the Moon, Giacconi *et al.* (1962) discovered the first extrasolar X-ray source, *Sco X-1*.

Various theories were developed to explain the puzzling nature of Sco X-1: its luminosity was millions of times higher than that of the Sun in the X-rays. Shklovsky (1967) proposed the idea that Sco X-1 could be a neutron star accreting matter from its environment. Following the era of rockets and balloons, soon X-ray astronomy would enter its observatory era with *Uhuru* (e.g., Giacconi *et al.* 1971) which surveyed the entire sky, discovering numerous X-ray sources, and confirming the existence of compact objects in X-ray sources. Furthermore, discovery of pulsations, and eclipses

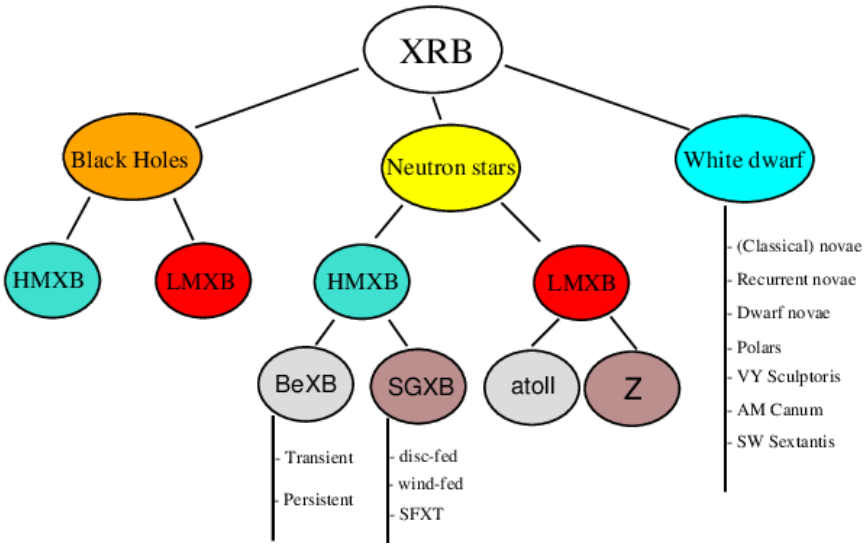


Figure 1.1: Classification of X-ray binaries (Reig 2011).

showed that these *X-ray stars*, are indeed close binaries, and it was the release of gravitational energy by material falling onto the compact object that is powering their X-ray emission.

We now know that these sources — called *X-ray binaries* (XRBs) nowadays — are composed of a compact object, i.e., white dwarf, neutron star, or black hole, which is accreting matter from a *companion* star. Hundreds of systems have been discovered in the Galaxy, its neighborhood (e.g., Magellanic Clouds), but also in distant galaxies.

The properties of XRBs depend on the nature, mass, and evolutionary stage of the stars composing the system, which in combination with its orbital parameters, determine the mass-transfer mechanism. These lead to a large variety in phenomenology which has been predominantly explained on the basis of different accretor and donor stars (see Figure 1.1). Accreting white dwarfs, give rise to a 'zoo' of different types of systems which we refer to as *cataclysmic variables*. Due to their relatively low luminosities ($\lesssim 10^{33}$ erg s^{-1}) compared to the other types of XRBs (e.g., Seward & Charles 2010), cataclysmic variables are typically not detected in extragalactic XRB surveys, and will not be discussed in the framework of this thesis. Henceforth, when referring to XRBs we are excluding cataclysmic variables as a sub-class.

The mass of the companion can affect the properties of XRBs in numerous ways. Firstly, a very massive star can have intense stellar winds, expelling material at a high rate. In a binary with a compact object, a fraction of this material will be captured by the

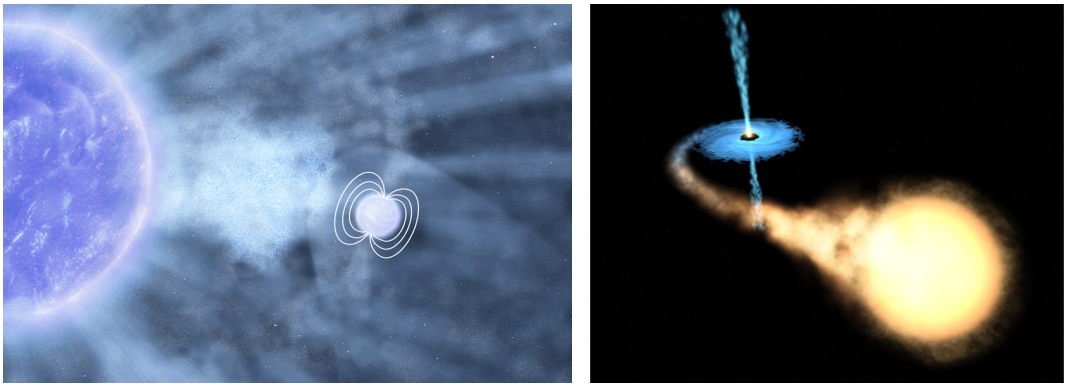


Figure 1.2: Artist's impressions on mass-transfer mechanisms. **Left**: an accreting binary system, consisting of a neutron star and a blue supergiant star. Intense stellar winds from the supergiant drive the accretion onto the compact object, releasing energy in the form of X-ray emission. 'Clumps in the wind' as the one depicted in the picture, may be responsible for sporadic flares last hours, as those observed in the case of Supergiant Fast X-ray Transients. *Credit: ESA/AOES Medialab.* **Right**: the GRO J1655-40, a system undergoing mass-transfer from an evolved star on to a black hole via Roche-lobe overflow, forming an accretion disk and a jet. *Credit: ESA, NASA, and F. Mirabel.*

compact object and, eventually, be accreted onto it. An example of this mass-transfer scenario is illustrated in the left panel of Figure 1.2.

On the other hand, stellar winds in low-mass stars are too weak to power an XRB. In the last stages of their evolution, however, the stars may fill their Roche lobe, forming a channel for material to flow from the star to the compact object through the L2 point. This mass-transfer mechanism is known as *Roche Lobe overflow* (RLO; see right panel of Figure 1.2). Note, however, that RLO is not characteristic of only low-mass stars; any star that during its evolution fills its Roche Lobe may transfer matter to its companion.

Observational studies indicate a 'bimodal' distribution in the donor mass. Specifically, the donor stars in most XRBs are either less massive than the Sun ($<1\,M_\odot$), i.e. low-mass XRBs (LMXBs), or significantly higher in mass ($>10\,M_\odot$), i.e. high-mass XRBs (HMXBs). Although intermediate-mass XRBs also exist (e.g., Podsiadlowski & Rappaport 2000), LMXBs and HMXBs dominate the population of XRBs. This behaviour may be the result of the complicated evolutionary paths of binary stars given their initial parameters (e.g., initial mass function, orbital separation distribution, mass ratio etc.)

In Figure 1.3 we show the evolution of two cases of progenitors of XRB systems, and use them as examples to explain some of the factors that come into play for their evolution into a LMXB and a HMXB. In the left panel we see the evolution of a LMXB system: after an initial Roche-lobe overflow before the formation of the compact object, the more massive star evolves into a red giant, leading to a *common envelope* phase which brings the two companions closer together. So close, that during the evolution of the second star, a second, very effective RLO phase begins, effectively powering a LMXB. In the right panel both stars are very massive initially. The initial Roche-lobe overflow transfers a large fraction from the primary star to the secondary, delaying the evolution of the former, and speeding up the evolution of the latter. The primary star, stripped from its outer layer's hydrogen, misses its giant phase, and the system does not pass a common envelope phase. After the first supernova explosion, the orbital system becomes too wide for a second RLO. However, the 'rejuvenated' secondary star undergoes mass loss due to intense winds. The 'loaned' material returns into its original owner which now is a neutron star. The hot accretion disk emits in the X-rays, and the system is a HMXB for about 10 Myr. Subsequent evolution of the secondary star leads short common-envelope and RLO phases.

Due to their complex evolutionary paths, XRBs have played an important role in understanding binary evolution, and are laboratories for constraining binary evolution models (e.g., Rappaport *et al.* 2005; O'Shaughnessy *et al.* 2008).

1.1.1 Demographics of X-ray binaries

Since high-mass stars have shorter lifetimes than low-mass stars, so do the companions of HMXBs. For this reason, the latter are typically found in young stellar populations, and therefore are associated to star-forming regions (e.g., Ho *et al.* 2001). We expect to find them in the disks of spiral galaxies, as well as irregular galaxies (see left panel of Figure 1.4). Elliptical galaxies, and dwarf spheroidal galaxies, having quenched their star-formation, are not expected to host HMXBs. On the other hand, the slow evolution of LMXBs — of the order of Gyr — means that their progenitors are formed at earlier stages of the cosmic history, when galaxies were building their stellar mass. Therefore, having lost their initial spatial distribution, they are found spread out according to the mass distribution of their host galaxies. Elliptical galaxies having large stellar masses, typically contain more LMXBs than spiral galaxies (e.g.,

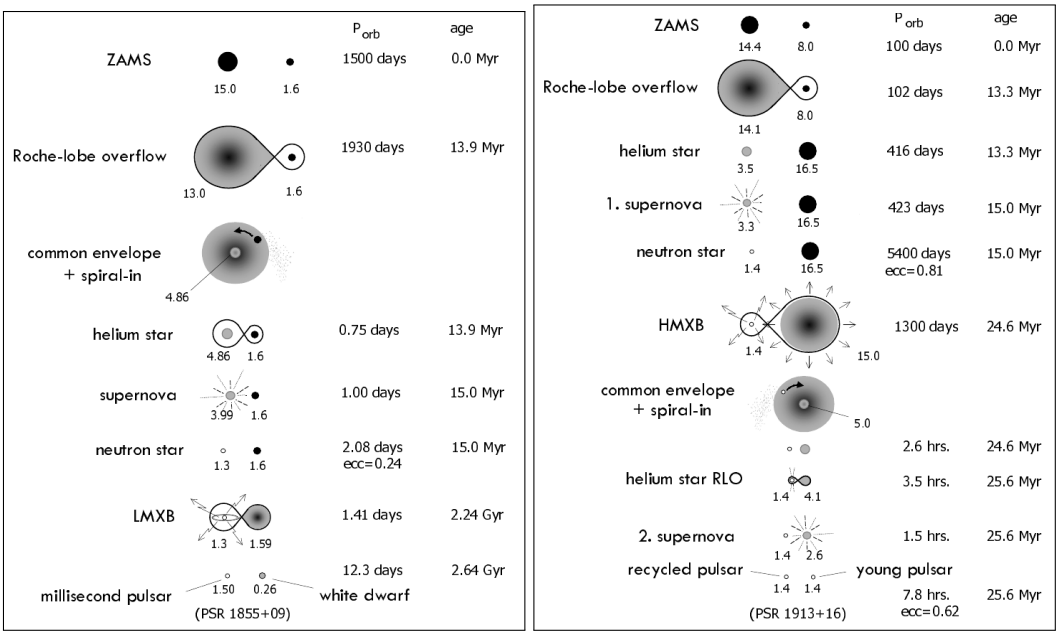


Figure 1.3: Common examples of evolution of XRB progenitor systems, a LMXB powered by Roche-lobe overflow (left), and a HMXB powered by wind accretion. *Credit: Tauris & van den Heuvel (2006)*

.

Gilfanov 2004). In addition, the dynamical formation of LMXBs in globular clusters (e.g., Clark 1975), increases the population of LMXBs in elliptical galaxies which, in general, host large globular cluster populations (see right panel of Figure 1.4).

The above trends are manifested in a correlation of the the number of LMXBs and HMXBs with the stellar mass and star-formation rate (SFR) of the host galaxy, respectively (e.g. Grimm *et al.* 2003; Gilfanov 2004). Therefore, SFR-scaled luminosity functions (LFs) for HMXBs and stellar-mass–scaled LFs of LMXBs, have been constructed using X-ray observations of samples of star-forming and elliptical galaxies, respectively (e.g., Grimm *et al.* 2003; Gilfanov 2004; Mineo *et al.* 2012; Zhang *et al.* 2012). Although the LFs agree in general, their exact numbers differ depending on the sample, luminosity range, etc.

For example, Mineo *et al.* (2012) found that a SFR-scaled, broken power-law, describes the luminosity function of HMXBs, based on a sample of galaxies with high

Figure 1.4: Examples of the spatial distribution of XRBs in star-forming and quenched galaxies. **Left**: image of the star-forming galaxy NGC2207 in the infrared (lower), X-rays (middle) and X-ray + IR + optical composite (upper). Being an actively star-forming galaxy, it hosts a large number of HMXBs which are spatially correlated with star-forming regions (as indicated in the IR image). *Credit: X-ray: NASA/CXC/SAO/S.Mineo, Optical: NASA/STScI, Infrared: NASA/JPL-Caltech.* **Right**: composite optical + X-ray image of NGC4649. The point sources follow a radially symmetric pattern, following the mass distribution of the elliptical galaxy, as well as, that of its globular cluster system. *Credit: X-ray: NASA/CXC/MSU/J.Strader, Optical: NASA/STScI.*

specific star-formation rate:

$$\frac{\mathrm{d}N}{\mathrm{d}L_{38}} = \xi\ \mathrm{SFR} \times \begin{cases} L_{38}^{-\gamma_1}, & L_{38} < L_\mathrm{b}, \\ L_\mathrm{b}^{\gamma_2-\gamma_1}\, L_{38}^{-\gamma_2}, & L_\mathrm{b} \le L_{38} \le L_\mathrm{cut}, \end{cases} \tag{1.1}$$

where L_{38}=$L_\mathrm{X}/10^{38}$ erg s^{-1}, L_b is the break luminosity, γ_1 and γ_2 are the power-law indices before and after the break, and L_cut is the cut-off luminosity, fixed at $5{\times}10^3$. The best-fit values were: ξ= (1.49±0.07) $\mathrm{M}_\odot^{-1}$ yr, γ_1=1.58±0.02, γ_2=$2.73^{+1.58}_{-0.54}$, and L_b=110^{+57}_{-34}. In the left panel of Figure 1.5 we reproduce figure 14 from Mineo *et al.* (2012) showing the cumulative form of the best-fitting LF, on top of the LFs of individual galaxies.

Similarly, based on a sample or early-type galaxies, Zhang *et al.* (2012) fitted a stellar-mass–scaled, doubly-broken power-law to the luminosity function of LMXBs

(introduced by Gilfanov 2004):

$$\frac{\mathrm{d}N}{\mathrm{d}L_{36}} = M_\star \times \begin{cases} K_1 \left(L_{36}/L_{\mathrm{b},1}\right)^{-a_1}, & L_{36} \leq L_{\mathrm{b},1}, \\ K_2 \left(L_{36}/L_{\mathrm{b},2}\right)^{-a_2}, & L_{\mathrm{b},1} \leq L_{36} < L_{\mathrm{b},1}, \\ K_3 \left(L_{36}/L_{\mathrm{cut}}\right)^{-a_3}, & L_{\mathrm{b},2} \leq L_{36} \leq L_{\mathrm{cut}}, \end{cases} \tag{1.2}$$

where L_{36}=$L_{\mathrm{X}}/10^{36}$ erg s^{-1}, $L_{\mathrm{b},1}$, $L_{\mathrm{b},2}$, and L_{cut} are the two luminosity breaks and the cut-off luminosity, respectively, and a_1, a_2, a_3 are the tree power-law indices below each break and the cut-off. The normalisation K_1 expresses the scaling with the stellar mass, while the normalisations, K_2 and K_3 are related to K_1 as follows:

$$\begin{aligned} K_2 &= K_1 \left(L_{\mathrm{b},1}/L_{\mathrm{b},2}\right)^{a_2}, \\ K_3 &= K_2 \left(L_{\mathrm{b},2}/L_{\mathrm{cut}}\right)^{a_3}. \end{aligned} \tag{1.3}$$

The best-fitting parameters where a_1=$1.02^{+0.07}_{-0.08}$, a_2=$2.06^{+0.06}_{-0.05}$, a_3=$3.63^{+0.67}_{-0.49}$, $L_{\mathrm{b},1}$=$54.6^{+4.3}_{-3.7}$, $L_{\mathrm{b},2}$=599^{+95}_{-67}, and K_1=1.01±0.28 per 10^{11} M$_\odot$ of stellar mass. The right panel of Figure 1.5 reproduces the fig 2 from Zhang *et al.* (2012) showing the cumulative LFs in early-type galaxies.

The above mentioned studies of LFs have been used to obtain scaling relations of the number of XRBs as a function of the stellar population parameters of the galaxies, namely their SFR and stellar mass:

$$N\left(> 5{\times}10^{37}\ \mathrm{erg\,s^{-1}}\right) = 5.4 \times M_\star \left(10^{10}\,\mathrm{M_\odot}\right), \tag{1.4}$$

$$N\left(> 10^{38}\ \mathrm{erg\,s^{-1}}\right) = 3.22 \times \mathrm{SFR}\left(\mathrm{M_\odot\,yr^{-1}}\right). \tag{1.5}$$

However, the samples used for such scaling relations are based on either highly star-forming galaxies (to exclude contribution from LMXBs) or early-type galaxies (to exclude contribution from HMXBs). For this reason, Lehmer *et al.* (2019) introduced an XRB LF which accounts simultaneously for the contribution of both LMXBs and HMXBs:

$$\frac{\mathrm{d}N}{\mathrm{d}L} = \frac{\mathrm{d}N_{\mathrm{HMXB}}}{\mathrm{d}L} + \frac{\mathrm{d}N_{\mathrm{LMXB}}}{\mathrm{d}L}, \tag{1.6}$$

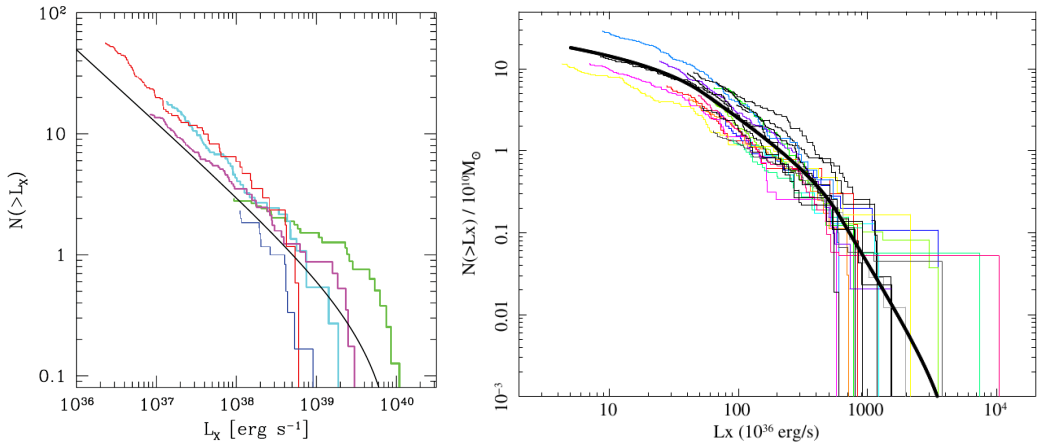


Figure 1.5: Cumulative luminosity functions of HMXBs and LMXBs in nearby galaxies. **Left**: The number of HMXBs per star-formation unit ($M_\odot$ yr^{-1}) as a function of the X-ray luminosity, in seven representative galaxies from the sample of Mineo *et al.* (2012). The solid line represents the best-fit HMXB LF in Mineo *et al.* (2012). **Right**: The number of LMXBs per 10^{10} $M_\odot$ of stellar mass in the sample of Zhang *et al.* (2012). The solid line is the best-fit LMXB LF in Gilfanov (2004).

which is the sum of a simple power-law HMXB LF, and a broken power-law LMXB LF.

Furthermore, the metal content (i.e., metallicity) of the star-forming gas in a galaxy can affect the number of HMXBs as indicated in observational (e.g., Mapelli *et al.* 2010; Basu-Zych *et al.* 2013a,b; Brorby *et al.* 2014; Douna *et al.* 2015) and theoretical studies (e.g., Linden *et al.* 2010; Fragos *et al.* 2013a,b). These studies showed that metal poor galaxies tend to have larger populations and/or more luminous X-ray sources. Recently, a metallicity dependent LF was measured by Lehmer *et al.* (2020), finding that different metallicities affect mostly the brightest sources (i.e. $>10^{38}$ erg s^{-1}).

Along with the metallicity effect, another important factor in the observed XRB populations is the effect of the age of the parent stellar populations. This indicated by binary population synthesis models (e.g., Fragos *et al.* 2008, 2013a). The effect of age in the formation efficiency of HMXBs has been observed in a study of HMXBs at sub-galactic scales in the Magellanic Clouds (Antoniou & Zezas 2016), and the evolution of the integrated X-ray luminosity of XRB populations as a function of age (Lehmer *et al.* 2017). These studies found that the populations of HMXBs and their luminosities peak at ~30 − 50 Myr, while at longer timescales (~Gyr) they decline. Similarly, an age-effect in LMXB LFs, is hinted in Lehmer *et al.* (2019).

1.2 Ultraluminous X-ray Sources

1.2.1 History

Studies of X-ray point source populations in external galaxies with the *Einstein* observatory, discovered off-nuclear sources (i.e. not associated with AGN) with luminosities exceeding 10^{39} erg s^{-1} (Long & van Speybroeck 1983; Fabbiano 1989). As the Eddington limit for an accreting compact object of mass M is $\sim 1.3\times10^{38} M/\mathrm{M}_{\odot}$ erg s^{-1}, the luminosities of these sources point at black holes more massive than those typically found by studies of black hole X-ray binaries in the Milky Way ($\lesssim 15\,\mathrm{M}_{\odot}$; e.g., Tanaka 1989; Remillard & McClintock 2006).

Among the various initial speculations for the origin of ULXs were background AGN, luminous young supernovæ, intermediate-mass black holes (IMBH; i.e., with masses in the $10^2 - 10^4\,\mathrm{M}_{\odot}$ range; Colbert & Mushotzky 1999), but were soon dismissed on the basis of their numbers, variability and spectral properties, lack of associations with radio sources, and problems in evolving IMBHs (e.g., Zezas & Fabbiano 2002; King 2004; van der Marel 2004; Rappaport *et al.* 2005; Gladstone *et al.* 2009).

The most viable scenario was that ULXs are accreting black-hole binaries undergoing near- or super-Eddington accretion, with non-isotropic emission due to geometrical beaming (e.g., Shakura & Sunyaev 1973; King *et al.* 2001; Begelman 2002; King 2009). The recent discovery of pulsar ULXs, i.e. ULXs with a neutron star accretors (and therefore significantly lower masses than black holes), proved that luminosities significantly higher than the Eddington limit are possible (e.g., Bachetti *et al.* 2014; Fürst *et al.* 2016; Israel *et al.* 2017a,b; Carpano *et al.* 2018). Furthermore, it definitively showed that neutron star ULXs may be a significant fraction of the ULX populations, challenging our understanding of ULX formation paths (e.g., King & Lasota 2016; King *et al.* 2017; Middleton & King 2017), and motivating more detailed binary population synthesis studies in order to account for neutron-star ULXs (e.g., Wiktorowicz *et al.* 2017; Misra *et al.* 2020).

1.2.2 Implications

The study of ULXs as extreme examples of compact X-ray sources is important for probing accretion physics at high mass-transfer rates, and indicating whether ULXs are characterised by a a new mode of accretion, the *ultraluminous state* (e.g., Gladstone

et al. 2009). In this, *supercritical accretion state*, the Eddington luminosity corresponding to a spherically accreting system is exceeded in the presence of a geometrically slim or thick accretion disk (Abramowicz *et al.* 1988). The physical processes governing the accretion disk (e.g., photon trapping: Begelman 1978; outflows: Shakura & Sunyaev 1973, formation of an inner funnel: Abramowicz *et al.* 1988) can provide useful insights on its geometry and spectral energy distribution, as well as the properties of sub, near and super-Eddington XRBs (e.g., Sądowski *et al.* 2014a). In addition, understanding accretion physics in stellar-mass binaries may extend our knowledge on the spectra, variability, and formation of other, more massive, accretion-powered sources such as IMBHs and quasi-stellar objects.

By constraining models for the formation and evolution of ULXs, we can advance our knowledge on stellar and binary evolution theory (e.g., Rappaport *et al.* 2005). The latter is crucial for the study of other manifestations and states of compact binary stars, gravitational-wave and gamma-ray burst sources (e.g., Abbott *et al.* 2017b), which furthermore may have passed through a ULX phase (e.g., Finke & Razzaque 2017; Marchant *et al.* 2017).

Observational studies indicate an excess in the total X-ray output of low-metallicity galaxies, suggesting an enhancement in the numbers and/or the luminosities of XRBs (e.g., Douna *et al.* 2015; Brorby *et al.* 2016). This metallicity trend could be the origin of a similar excess observed in high-redshift galaxies since they typically present lower metallicities (e.g., Lehmer *et al.* 2005; Basu-Zych *et al.* 2013a,b; Fornasini *et al.* 2020). This has implications for the contribution of XRBs to the heating of the intergalactic medium (IGM) before and during the epoch of reionisation (e.g., Madau *et al.* 2004; Kaaret 2014; Pacucci *et al.* 2014). While initial studies indicated negligible effects from XRBs to the heating of the IGM (e.g., Madau & Fragos 2017), the few but bright ULXs may ionise the interstellar medium of the host galaxies, lowering the absorption in the soft X-rays, ultimately allowing them to penetrate the host galaxy and heat the IGM even more (e.g., Sazonov & Khabibullin 2018). Quantifying whether the metallicity elevates the number of XRBs in general, or boosts their luminosities, is critical for our knowledge of the X-ray luminosity and spectral energy distribution of the galaxies, and its imprint in the thermal history of the Universe.

1.2.3 ULX populations in the context of their host galaxies

The dismissal of the initial scenarios for the origin of ULXs, led to the conclusion that ULXs are objects in the high-end of the HMXB LF. The connection of ULXs with HMXBs is supported by their direct association with star-forming regions in individual galaxies, such as the Antenæ galaxies: (e.g., Fabbiano *et al.* 2001; Zezas *et al.* 2007), the Cartwheel galaxy (e.g., Gao *et al.* 2003; Wolter & Trinchieri 2004), NGC4485/4490 (e.g., Roberts *et al.* 2002), and Arp299 (e.g., Zezas *et al.* 2003; Anastasopoulou *et al.* 2016). This motivated demographic studies of large samples of ULXs and their host galaxies, which estimated the scaling of the number of ULXs with the SFR of their host galaxy (e.g., Swartz *et al.* 2004, 2011; Luangtip *et al.* 2015). For example, Swartz *et al.* (2011) found ~2 ULXs per $M_{\odot}\ yr^{-1}$. However, the HMXB LF of Mineo *et al.* (2012) predicts ~0.6 ULXs per $M_{\odot}\ yr^{-1}$. This tension may be rooted in the different host galaxy samples used in different surveys.

Interestingly, ULXs are also found in elliptical galaxies, but due to the initially small galaxy samples and large background source contamination, early surveys could not reliably quantify their scaling with the stellar mass (e.g., Angelini *et al.* 2001; Colbert & Ptak 2002; Irwin *et al.* 2004; Swartz *et al.* 2004). Constraints of the ULX-stellar mass scaling factor (e.g., Plotkin *et al.* 2014; Wang *et al.* 2016) are consistent with a specific ULX frequency of ~6 ULXs per $10^{12} M_{\odot}$. This value is in agreement with the LMXB LF of Zhang *et al.* (2012), suggesting that ULXs in elliptical galaxies contribute to the high-luminosity tail of the LMXB LFs. This indicates that donors of ULX systems can also be of late-type (e.g., Feng & Kaaret 2008; Wiktorowicz *et al.* 2017). However, this agreement is based on host galaxy samples at a narrow range in the stellar mass (massive elliptical galaxies).

Surprisingly, dwarf star-forming galaxies present an excess of ULXs with respect to their SFR (e.g., Swartz *et al.* 2008; Walton *et al.* 2011; Plotkin *et al.* 2014; Tzanavaris *et al.* 2016). This has been attributed to their lower metallicities, and it is highlighted by the observed excess of ULXs in metal-poor galaxies (e.g., Soria *et al.* 2005; Prestwich *et al.* 2013; Mapelli *et al.* 2010). These trends are similar to those observed in normal XRBs (e.g., Brorby *et al.* 2014; Douna *et al.* 2015), providing further support to the notion that ULXs are the bright-end of the XRB populations. The metallicity effect in the number of ULXs is mainly attributed to the weaker stellar winds in the low-metallicity progenitors. The first consequence of the latter is that the black holes retain significant

fraction of their initial mass (e.g., Heger *et al.* 2003; Belczynski *et al.* 2010), resulting into more massive accretors in black-hole XRBs which exhibit higher efficiency in converting gravitational energy to X-ray emission. The second consequence is that the angular momentum losses are small, leading to tighter orbits which favour the more efficient Roche-lobe overflow mass transfer, and hence the fraction of binaries that evolve into luminous XRBs (e.g., Linden *et al.* 2010).

1.3 Aims of this study

The purpose of this thesis is to study in detail the populations of ULXs in the context of the host galaxy properties such as star-formation rate, stellar mass, and metallicity. The results will provide up-to-date and tight constraints for binary population synthesis models, and aid in the design of new observational campaigns.

In order to perform this analysis we require large samples of host galaxies and X-ray sources, and to overcome a number of challenges:

I. ULXs are rare (typically 0-2 per galaxy), and therefore are usually found at distant galaxies.

II. At large distances, sources may be blended together, which can bias both their numbers and their luminosity distribution.

III. Contamination from foreground stars and background AGN, especially in nearby, extended galaxies.

IV. The required host galaxy parameters (e.g., distance, SFR, stellar mass, metallicity), are not readily available for large samples of galaxies.

V. Nuclear activity in the host galaxy may bias the estimated stellar population parameters.

Addressing these challenges motivated three stand-alone, albeit interconnected, projects which are presented in three individual chapters:

1.3.1 The Heraklion Extragalactic Catalogue — *Chapter 2*

The first step is to define the galaxy sample that will be used for the census of ULXs. Since there is no complete galaxy catalogue providing the required data for our purpose, we compile a new galaxy compilation, the *Heraklion Extragalactic Catalogue* (*HECATE*).

Being based on the most complete extragalactic database, the *HyperLEDA* (Makarov *et al.* 2014), the *HECATE* includes all known galaxies in the local Universe, and hence all potential hosts of sources detected from X-ray surveys (addressing challenge **I**). Since the primary motivation for building the catalogue is the study of X-ray binaries in nearby galaxies, we set a maximum distance of 200 Mpc, well above the distance at which ULX demographics are possible (because of challenge **II**).

By incorporating data from extragalactic databases, and multi-wavelength surveys, this new catalogue provides

1. positions and angular sizes of galaxies, which can be used to associate X-ray sources;
2. robust distances;
3. morphological classifications which are necessary for revisiting demographic results for different types of galaxies;
4. stellar population parameters that can be used for the study of the ULX scaling relations (challenge **IV**); and
5. nuclear activity classifications (challenge **V**).

The *HECATE*, being an all-sky catalogue of all known galaxies in the local Universe, can also be useful for a wide range of applications, beyond the study of XRBs. The data provided by the catalogue can be used for the initial characterisation of transient astrophysical sources (e.g., supernovæ, tidal disruption events, gamma-ray bursts etc.) In addition, in the case of the poorly-localised gravitational wave sources, follow-up surveys of their electromagnetic counterparts can be aided by the prioritisation of potential host galaxies based on their properties in the *HECATE*.

1.3.2 A census of ULXs in the local Universe — *Chapter 3*

In this chapter, we associate the *HECATE* with X-ray sources. In order to address the issue of source blending (challenge **II**), we use data from the high-resolution X-ray observatory, *Chandra*, which allows us to probe ULX populations up to a distance of ~40 Mpc. Using the new *Chandra Source Catalogue 2.0* (*CSC 2.0*), we probe the X-ray source populations in thousands of galaxies.

However, the classification of X-ray sources based on optical associations or timing and spectral analysis is not feasible for the majority of the sources in our sample. Therefore, interlopers (challenge **III**) are removed statistically.

The combination of the *HECATE* and the *CSC 2.0* and the tools developed in this work, results in the most comprehensive demographic study of ULXs in the context of the stellar population parameters of the host galaxies.

1.3.3 Source confusion — *Chapter 4*

Since the resolution of *Chandra* cannot be matched by any current, or upcoming X-ray observatory, the problem of source blending (or confusion; challenge II) will continue to limit the studies of XRB populations. In this chapter, we quantify the effects of source confusion in the demographics of XRBs, and luminosity functions, as a function of distance, and we find the optimal distance that maximises the galaxy samples, while ensuring manageable confusion.

Finally, in Chapter 5 we present a summary of the results, and the conclusions drawn from them, as well as future extensions of this work.

2

The Heraklion Extragalactic Catalogue

2.1 Background

With the availability of all-sky surveys across the electromagnetic spectrum (e.g., *LSST*, *ZTF*, *eROSITA*) and the advent of the era of multi-messenger observations (e.g., gravitational-wave, neutrino, cosmic-ray observatories) there is an increasing need for homogenised extragalactic catalogues that can be used for the characterisation of individual sources.

Astronomical databases like *NED* (Helou *et al.* 1991), *SIMBAD* (Wenger *et al.* 2000) and *HyperLEDA* (Makarov *et al.* 2014) have significantly boosted extragalactic research via the collection, and organisation of data such as positions, distances, photometric fluxes, and morphological classifications. However, due to the diversity of the different sources of these data, they cannot be readily used for studies requiring derived galaxy properties such as star-formation rate (SFR), stellar mass ($M_\star$), metallicity, and nuclear activity, for large samples of objects. Although detailed catalogues based on focused surveys provide such information (e.g., *MPA-JHU*; Kauffmann *et al.* 2003; Brinchmann *et al.* 2004; Tremonti *et al.* 2004), the lack of all-sky coverage limits their usefulness

for many astrophysical applications, such as characterisation of sources in multi-wavelength all-sky or serendipitous surveys (e.g., X-ray surveys; Kim *et al.* 2007; Saxton *et al.* 2008).

The rapid identification of counterparts of transient sources such as gamma-ray bursts (GRBs) or rare events (e.g., high-energy cosmic rays), and the strategic planning of follow-up observations, are possible with the aid of all-sky galaxy catalogues. Furthermore, the use of astrophysical information has been used to increase the effectiveness of identifying the hosts of gravitational-wave (GW) sources (e.g., Abbott *et al.* 2017b). To this extent there is a growing effort to build galaxy catalogues providing information on $M_\star$ or SFR (or their proxies) aiming to aid GW follow-up observations (e.g., Kopparapu *et al.* 2008; White *et al.* 2011; Gehrels *et al.* 2016; Cook *et al.* 2019; Dálya *et al.* 2018; Ducoin *et al.* 2020). However, these catalogues do not provide metallicity, which can be a key factor for the identification of GW hosts (e.g., Artale *et al.* 2019). Since the aforementioned galaxy catalogues were designed for applications focusing on distant galaxies (e.g., GWs, GRBs), the provided data may not be very accurate for nearby galaxies (e.g., $D<40$ Mpc), which often require special treatment (e.g., extended vs. point-source photometry, and distance measurements vs. application of the Hubble-Lemaître law). Therefore, studies involving nearby galaxy samples often invest in compiling the necessary galaxy data from scratch.

In order to enable large-scale studies of transient events such as those described earlier, or multi-wavelength properties of galaxies (e.g., X-ray or γ-ray scaling relations; Ackermann *et al.* 2012; Komis *et al.* 2019; Kovlakas *et al.* 2020), we require an all-sky catalogue that gives accurate locations, galaxy dimensions, distances, multi-band photometry, and most importantly derived stellar population parameters. For this reason, we compiled an all-sky value-added catalogue of 204,733 nearby galaxies within a distance of 200 Mpc: the *Heraklion Extragalactic Catalogue* (*HECATE*[*]). This catalogue provides all the aforementioned quantities based on a variety of sources. Special care is taken to develop procedures that consolidate the available data, maximise the coverage of the parameters, and address possible biases and offsets between different parent catalogues. The derivation of homogenised stellar population parameters, including metallicity, and nuclear activity classifications, highlight the usefulness of the *HECATE* as a reference sample for the characterisation of sources in multi-wavelength and/or

[*] *Hekátē* (greek, Εκάτη), goddess of crossroads and witchcraft in ancient Greek mythology. Pronunciation: *hek-UH-tee*.

multi-messenger observations. The catalogue is publicly available at the *HECATE* portal: https://www.ia.forth.gr/hecate.

In §2.2 we describe the selection of galaxies from the *HyperLEDA* database, and the incorporation of redshift (z) and size information. The assembly and combination of distance measurements, as well as the derivation of z-dependent distances for galaxies without distance measurements is described in §2.3. The compilation of multi-wavelength data and the derivation of stellar population parameters is presented in §2.4. In §2.5 we compare the *HECATE* with other galaxy catalogues, discuss its limitations, and present various applications. Finally, in §2.6 we present future extensions of the catalogue. Throughout the paper, unless stated otherwise, uncertainties correspond to 68% confidence intervals.

2.2 Sample selection

As the basis of our catalogue, we use the *HyperLEDA* database (Makarov *et al.* 2014), which includes, combines and homogenises extragalactic data in the literature, without explicit flux or volume limits. Furthermore, common problems such as misprints, duplication, poor astrometry and wrong associations that can be found in legacy catalogues (e.g., *UGC*: Nilson 1973, *RC3*: de Vaucouleurs 1991) are generally identified and rectified by the *HyperLEDA* pipeline.

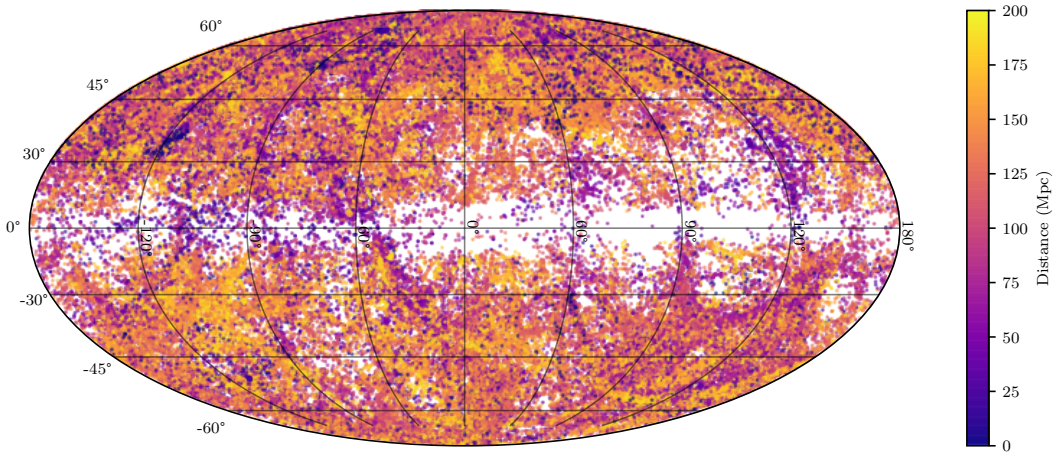


Figure 2.1: Sky map of the galaxies in the *HECATE* in Galactic coordinates, colour-coded according to their distance.

Out of the 5,377,544 objects in the *HyperLEDA* (as of October 2019), we select 3,446,810 (64%) that are characterised as individual galaxies ('objtype=G'), excluding groups, clusters, parts of galaxies, stars, nebulæ etc.

Since the distances for the majority of the galaxies have not been measured, we perform the selection of local Universe galaxies based on a recession velocity limit. We note that reported heliocentric radial velocity measurements typically contain the components of the peculiar motions of the Sun and the Milky Way. The peculiar motions of the galaxies are generally not known. We correct for those of the Sun with respect to the local Universe by computing the Virgo-infall corrected radial velocity, $v_{\rm vir}$, which corrects for all motions of the Sun, and Milky Way up to the level of the infall of the Local Group to the Virgo Cluster. We select all galaxies with $v_{\rm vir}$<14, 000 km s^{-1} (corresponding to z<0.047 and $D\lesssim$ 200 Mpc assuming Hubble parameter h=0.7). The Virgo-infall corrected radial velocity in *HyperLEDA* is outdated (D. Makarov, private communication). Therefore, we compute it for all galaxies (see Appendix 2.A for details on the computation).

204,467 objects are found in *HyperLEDA* with $v_{\rm vir}$<14, 000 km s^{-1} while 2,560,816 exceeded the velocity limit. However, for the 681,527 galaxies in *HyperLEDA* without radial velocity measurements, we attempted to get measurements from *NED*. The association to *NED* is described in Appendix 2.B.1. In total, we recover the radial velocities for 1,494 additional objects with $v_{\rm vir}$<14, 000 km s^{-1}.

Note that in the above procedures, we performed various manual inspections to exclude duplicates in *HyperLEDA* or misclassified objects (stars, artefacts from diffraction light, 'parts of galaxies', etc.) In total 1,228 objects were rejected in this process. The final sample consists of 204,733 galaxies. Figure 2.1 shows a sky map of the *HECATE*.

Out of the 204,733 galaxies in our sample, there are 39,251 objects without size information, restricting the cross-matching capabilities of our sample. For the majority of them, the semi-major axis is complemented via cross-linking of our sample with other databases and surveys, resulting into 199,895 galaxies with size information (97.6%). The procedure is described in detail in Appendix 2.B.2.

Finally, in the *HECATE* we include additional information from *HyperLEDA* such as: astrometric precision, object name, morphological classification, optical photometry, inclination, and Galactic absorption. A full list of the information provided in the *HECATE* is given in Appendix 2.D.

2.3 Distance estimates

Robust distance estimates for the galaxies in the *HECATE* are essential for the purposes of this catalogue, and required for estimating the stellar population parameters of the galaxies.

While redshift-derived distances can be calculated using the Hubble-Lemaître law for the majority of the galaxies in the *HECATE* (positive z), this approach is not accurate in the case of nearby galaxies for which recessional velocities are dominated by their peculiar motions. In addition, this method cannot be used for blue-shifted galaxies*. Furthermore, at the distance range of the *HECATE*, the unknown peculiar motion of a given galaxy adds to the uncertainty on its distance, equally or more than the propagation of the uncertainties of the galaxy's z and the Hubble parameter. For this reason, we use z-independent distance measurements from *NED-D* where available (for ≈10% of the galaxy sample), and combine them with the method described in §2.3.1. For the remainder of the galaxies (≈90%), we estimate the distance of the galaxies using a regression method, described in §2.3.2, based on the sample of galaxies with known distances.

2.3.1 Redshift-independent distances

The largest resource of z-independent distances is the *NED-D* compilation, containing 326,850 measurements (as of March 2020) for 183,062 objects, based on 96 different distance indicators (Steer *et al.* 2017). However, for objects with multiple measurements, *NED-D* does not readily provide a summary of these distance estimates. On the other hand, *CosmicFlows 3.0* (*CF3*; Tully *et al.* 2016), reports distance estimates for 17,669 galaxies at $z \lesssim 0.05$, calculated as uncertainty-weighted averages of individual measurements. Aiming at an as-large-as-possible sample of galaxies with distance determinations, we obtain distance measurements from *NED-D* in order to combine them into unique estimates for each galaxy, and use the *CF3* for consistency checks.

We reject measurements that are not based on peer-reviewed sources, and those using outdated distance moduli for the Large Magellanic Cloud (i.e. outside the 18.3−18.7 mag range; Pietrzyński *et al.* 2013) or distance scales calibrated for Hubble

* In fact, the most blue-shifted galaxy in our sample with a reliable distance measurement is a Virgo Cluster member, VCC 0815, at distance of 19.8 Mpc, which corresponds to a recession velocity ∼1400 km s^{-1} but its heliocentric radial velocity is −700±50 km s^{-1}.

Table 2.1: List of distance indicators used in *HECATE* (see Steer *et al.* 2017 and references therein), the number of galaxies ($N_{\rm gal}$) for which the measurements (number $N_{\rm meas}$) based on each indicator were considered in the final distance estimates, and the corresponding typical uncertainty of the distance moduli, $\langle\sigma_\mu\rangle$, in mag.

Distance indicator	$N_{\rm gal}$	$N_{\rm meas}$ *	$\langle\sigma_\mu\rangle$
Cepheids	75	1416	0.11
Eclipsing binary	4	45	0.09
Fundamental plane	10697	26975	0.35
Horizontal branch	29	65	0.10
Red clump	14	102	0.09
RR Lyræ	38	282	0.09
Sosies	280	280	0.29
Surface brightness fluctuations	482	1650	0.18
Tip of the red giant branch	358	1361	0.13
Tully-Fisher	10780	11309	0.40

* We note that $N_{\rm meas} \geq N_{\rm gal}$ because for many galaxies there are multiple distance measurements based on the same indicator.

constants outside the range 60−80 km s^{-1} Mpc^{-1}. Many of the 93 distance indicators reported in *NED-D* are appropriate for objects at distances greater than the volume limit of the *HECATE* (e.g., SNIa) and therefore we do not consider them. We also avoid methods applied in less than three publications, as their systematic uncertainties or validity may be insufficiently understood. To be conservative, we select 10 commonly used indicators that are considered relatively reliable at distances <200 Mpc (e.g., Steer *et al.* 2017), listed in Table 2.1. For publications reporting for the same galaxy multiple individual measurements based on the same indicator (e.g., Cepheid distances for different stars within a galaxy), we adopt the concluding measurement in each publication. Reported zero distance uncertainties (10 cases) were treated as undefined. Preference is given to measurements with reported uncertainties over those without uncertainties. In total, we associate 43,511 distance measurements with 21,174 galaxies in the *HECATE*.

For the 13,247 galaxies with single distance measurements, we adopt them as they are, 97% of which have reported uncertainties. For the 591 galaxies with multiple measurements but no uncertainties, we use their weighted mean as the final estimate, and their weighted standard deviation as the uncertainty. The weights depend on the year of publication (penalising old measurements) to reduce historical biases (e.g.,

older calibrations, unknown biases):

$$w_i = \delta^{y_i - y_{\rm ref}}, \tag{2.1}$$

where δ is the penalty per year – we set δ=$2^{0.1}$ so that the weight is halved for every decade passed*, y_i is the year of measurement, and $y_{\rm ref}$ is an arbitrary reference year. We note that for seven galaxies out of these 591, the standard deviation was 0 (possibly duplicate measurements), and therefore we do not report the uncertainty of the final distance estimate.

For the 7,336 galaxies with multiple distance measurements and uncertainties, we calculate the final distances and corresponding uncertainties using a weighted Gaussian Mixture (GM) model. The weight for the i-th measurement is

$$w_i = \delta^{y_i - y_{\rm ref}} \sigma_i^{-2}, \tag{2.2}$$

where the parameters are the same as in Equation 2.1, with the additional contribution of the measurement uncertainty term σ_i. The GM distribution of the distance modulus μ of a galaxy is derived by combining M individual measurements:

$$f_{\rm GM}(\mu) = \sum_{i=1}^{M} w_i f_i(\mu) \Big/ \sum_{i=1}^{M} w_i, \tag{2.3}$$

where w_i are the weights calculated in Equation 2.2, and f_i is the PDF of the distribution of each measurement. We consider each measurement to be Gaussian-distributed, with mean and standard deviation equal to the distance modulus and its uncertainty reported in *NED-D*. We note that the mean of the distribution resulting from Equation 2.3 is mathematically equivalent to the weighted average of the individual means (and therefore consistent with the methods for galaxies with single measurement, or without uncertainties), while its spread accounts for both the dispersion of the measurements and their uncertainties.

* We chose this value because: (i) we found systematic offsets (0.05–0.2 mag) in the distance moduli measured at times with differences >20yr, and (ii) for small values of δ (corresponding to $^1/_2$-folding time-scales of less than 5 years), we found an increased scatter (>0.1 mag) because, effectively only few newer measurements contribute to the distance.

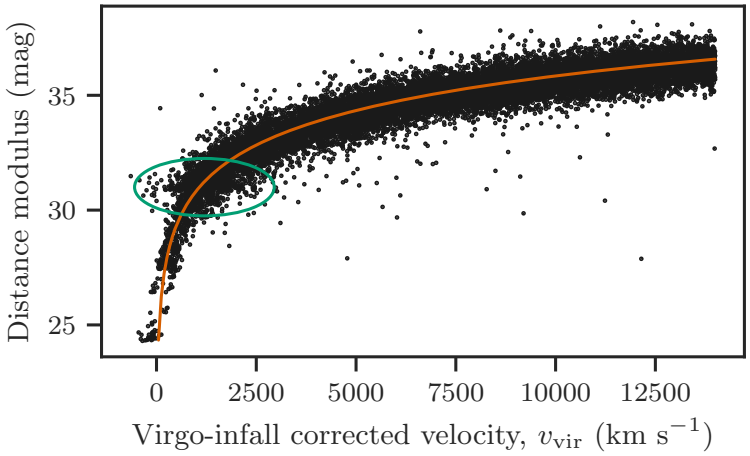


Figure 2.2: Hubble diagram of the galaxies in the *HECATE* with z-independent distances. The orange line shows the Hubble-Lemaître law with H_0=(67.8±0.1) km s^{-1} Mpc^{-1} (Planck Collaboration *et al.* 2016). We note that the majority of the points are following the law, albeit with significant dispersion at low values of $v_{\rm vir}$, and the existence of a 'branch' at distance modulus ~31 (indicated by a green ellipse) caused by Virgo Cluster galaxies which present significant velocity dispersion (cf. fig. 10 in Tully *et al.* 2016).

2.3.2 Redshift-dependent distances

For galaxies without distance measurements (≃90%), we rely on the spectroscopic redshift information. While we could simply use the Hubble-Lemaître law and the redshift of each object in order to calculate their distances, the proximity of the galaxies in the *HECATE* sample makes them very sensitive to peculiar velocities and local deviations from the Hubble flow. For this reason we adopt a data-driven approach where the galaxies with z-independent distances (10% of the full sample) are used as the training data set in a regression model that infers the distances (and uncertainties) at similar recession velocities for the rest of the sample. The uncertainties of the radial velocities were not accounted for in the regression since, in the case of spectroscopic redshifts, they are negligible compared to the uncertainties of the distance measurements in the training sample.

Figure 2.2 shows the distance modulus as a function of the radial velocity for the galaxies with z-independent distances in our sample (calculated as described in §2.3.1). We observe (i) that – unsurprisingly – the distance correlates with radial velocity even for nearby and blueshifted galaxies, albeit with higher dispersion, and (ii) a horizontal branch at distance modulus ≈ 31 mag that is caused by Virgo Cluster galaxies. In order

to account for such local deviations from the Hubble flow we employ a data-driven approach for robust distance estimates as follows:

1. the galaxy sample is separated into two subsamples: galaxies in the Virgo Cluster and the rest (§2.3.2.1);

2. for each subsample, a regressor is trained using the galaxies with redshifts and z-independent distances, so that the distance and its uncertainty are predicted from the recession velocity;

3. the distance and its uncertainty for the galaxies without z-independent measurements, is predicted using the two regressors.

2.3.2.1 Virgo-cluster membership

As we discussed in the previous paragraphs, and shown in Figure 2.2, special treatment of Virgo Cluster members is necessary for estimating their distance from the recession velocity. The most up-to-date catalogue of galaxies of the Virgo Cluster, the *Extended Virgo Cluster Catalog (EVCC)*, was produced by Kim *et al.* (2014) using the radial velocities and a cluster infall model, as well as morphological and spectroscopic classification schemes. The EVCC is cross-matched with our sample to identify the galaxies associated with the Virgo cluster.

2.3.2.2 Local average and standard deviation of Hubble diagram

We use the Kernel Regression technique (Nadaraya 1964) to compute the intrinsic distance modulus $\mu_{\rm int}(v_{\rm vir})$ at a given Virgo-infall corrected radial velocity, $u{\equiv}v_{\rm vir}$. The sample is split into Virgo members (VC) and non-Virgo members (nVC). For each subsample, we compute the local (at u) distance modulus, $\mu_{\rm int}(u)$ as the weighted average of the distance moduli μ_i of the N galaxies it contains, with weights $(q_i(u))$ given by the Guassian kernel with bandwidth h (or 'averaging length'). Similarly, for each subsample (VC and nVC) we calculate the 'local standard deviation' in terms of the bias-corrected weighted standard deviation:

$$\sigma_i(v) = \sqrt{\frac{V_1}{V_1^2 - V_2} \sum_{i=1}^{N} q_i(u)\,[\mu_i - \mu_{\rm int}(u)]^2}, \tag{2.4}$$

where $V_1 = \sum_{i=1}^{N} q_i(u)$ and $V_2 = \sum_{i=1}^{N} q_i^2(u)$.

The choice of the bandwidth h effectively sets the resolution, in radial velocity, of the derived statistical quantities. Due to the significant curvature of the radial velocity vs. distance modulus diagram (Figure 2.2) for $u \lesssim 2000\,\mathrm{km\,s^{-1}}$, the resolution should be increased in this region in order to capture the shape and prevent mixing of data from regions of significantly different slopes. On the other hand, at greater radial velocities (or distances) a relatively large bandwidth would allow more data points to contribute, and hence provide an estimate that is less influenced by outliers. For these reasons, we set the bandwidth h for the Gaussian Kernel to be a function of the radial velocity, increasing with radial velocity but also kept constant in the steep part of the diagram by enforcing a minimum value, $h_{\rm min}$:

$$h(u) = h_{\rm min} \times \max\left\{1, u/2000\,\mathrm{km\,s^{-1}}\right\}, \tag{2.5}$$

where $h(u)$ is the bandwidth of the Gaussian Kernel for points evaluated at Virgo-infall corrected radial velocity u. For the Virgo Cluster model, we keep the bandwidth fixed as the distance modulus is expected to be roughly the same, rendering such considerations irrelevant.

The baseline of the bandwidth, $h_{\rm min}$, must be chosen carefully as it can easily result into 'overfitting' if too small (only a few of the data points are considered for the fit in each bin), or, 'underfitting' if too large (introducing *lack-of-fit* variance). We find the optimal bandwidth, by minimising the total regression error $\mathcal{S}$, i.e. the quadratic sum of the regression errors, $\mathcal{S}_i$, corresponding to each galaxy. $\mathcal{S}_i$ is evaluated by employing the leave-one-out cross-validation technique: the i-th galaxy is removed from the sample, and its distance is estimated using the kernel regression. The residual between the true distance modulus, and the regressed one is $\mathcal{S}_i$. Additionally, when the optimal bandwidth has been found, we remove outliers based on the true distance modulus and the predicted one (and its uncertainty), by performing sigma clipping at the 3σ-level, and re-optimise for $h_{\rm min}$ iteratively until no outlier is found. We applied the above procedure and found optimal minimum bandwidth $h_{\rm min} = 68.2\,\mathrm{km\,s^{-1}}$ for non-Virgo galaxies, after removing 149 outliers (<1% of the nVC subsample). For the Virgo galaxies, the optimal bandwidth was $h_{\rm VC}$=294.5 $\mathrm{km\,s^{-1}}$, while only one outlier was found (<1% of the VC subsample).

2.3.2.3 Local intrinsic dispersion

The local standard deviation we compute in Equation 2.4 encompasses the uncertainties of the distance moduli due to measurement uncertainties and the *intrinsic scatter* of the true distance modulus. The latter is attributed to the peculiar velocities of the galaxies and the systematic uncertainties due to the distance ladder calibration. Given the model $\mu_{\rm int}(v)$ and following Kelly (2007) we formulate the error model

$$\mu_i = \mu_{\rm int}(v_i) + \epsilon_i + \epsilon_{\rm int}(v_i), \tag{2.6}$$

where ϵ_i is a Gaussian-distributed random variate with mean equal to 0 and standard deviation equal to the distance modulus uncertainty of the i-th galaxy, $\mu_{\rm int}$ is the local average and $\epsilon_{\rm int}(v_i)$ is a Gaussian-distributed variate with mean equal to 0 and standard deviation $\sigma_{\rm int}(v)$, which is a *local intrinsic scatter* model. We apply a Maximum Likelihood Estimator (MLE) to calculate the local intrinsic scatter, $\epsilon_i(v)$. We note that the uncertainties on radial velocities have not been considered in our analysis as they are typically one order of magnitude smaller ($\sim$10 km s^{-1}) than the optimal kernel bandwidth for both VC and nVC models ($\sim$100 km s^{-1}) and the typical peculiar velocities of galaxies ($\sim$100 km s^{-1}; e.g., Hawkins *et al.* 2003).

We apply the above Kernel Regression model to 617 Virgo galaxies and 182,326 nVC galaxies to derive their distances and uncertainties. Also, for 37 Virgo galaxies and 317 nVC galaxies with z-independent distances but no uncertainties, we apply the local intrinsic scatter model to estimate their uncertainty. We ensure that the two models are applied only to galaxies with radial velocities covered by the training data sets: $v_{\rm vir} \in \left[-792\,{\rm km\,s^{-1}}, 2764\,{\rm km\,s^{-1}}\right]$ for VC and $v_{\rm vir} \in \left[-481\,{\rm km\,s^{-1}}, 14,033\,{\rm km\,s^{-1}}\right]$ for nVC in order to avoid extrapolation (note that the ranges are expanded by half optimal bandwidth, $h_{\rm min}$), leaving only 12 objects in the *HECATE* without distance estimates.

For quick reference, in Appendix 2.C we provide empirical formulæ for the distance modulus of a galaxy $\mu_{\rm int}$, and its uncertainty $\epsilon_{\rm int}$, given its Virgo-infall corrected velocity, based on the results of the aforementioned methods.

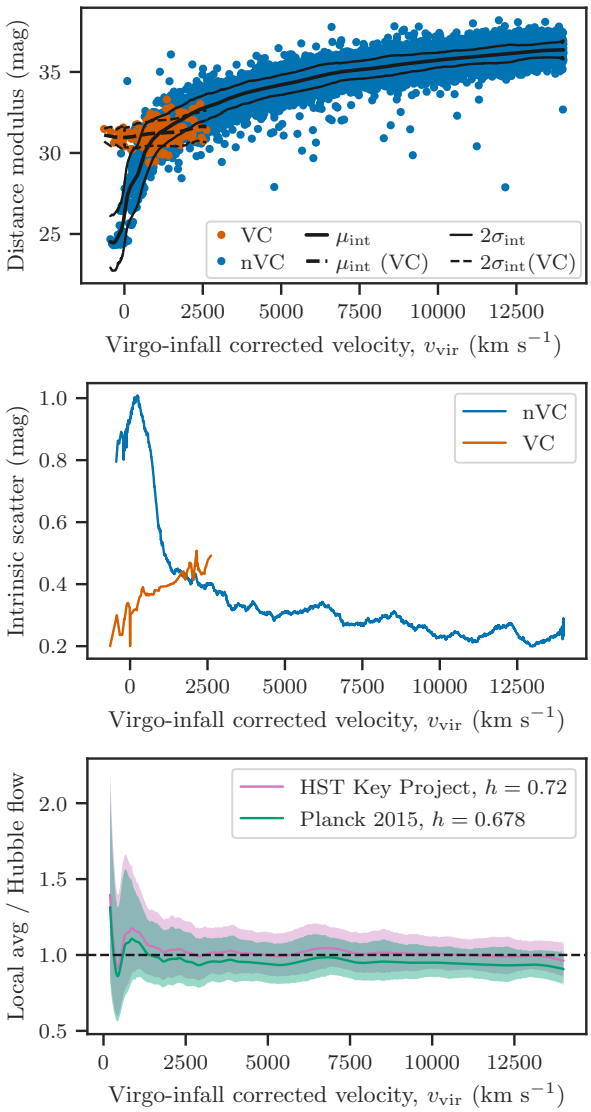


Figure 2.3: Assessment of the accuracy of the Kernel Regression models. The models capture the trends in the D-z in the local Universe, and provide accurate distances (0.2-0.4 mag or 10-20%) particularly for $v_{\rm vir} \gtrsim 2500$ ($D \gtrsim 35$ Mpc). **Top:** z-independent distances in our sample (points), separated to Virgo Cluster members (orange) and nVC galaxies (blue). The black lines depict the local mean and the 2-sigma regions (using the local standard deviation) according to the two regressors (dashed for VC, and continuous for nVC). **Middle:** The local standard deviation of the distance modulus of the two models. **Bottom:** The ratio of the local average distance (§2.3.2.2) and the distance inferred from Hubble-Lemaître law for two values of Hubble parameter: 0.72 (HST Key Project; magenta line) and 0.678 (*Planck* 2015; green line). For each ratio, we plot with the same colours, the 68% confidence region that reflects the local intrinsic scatter (§2.3.2.3).

2.3.2.4 Validation of the regression technique

The resulting distances from the Kernel Regression technique described above should reflect the trends of the z-independent distances used in the *HECATE*, and converge to the Hubble-Lemaître law for large distances.

The top panel of Figure 2.3 shows the distance moduli as a function of the radial velocity of the two subsamples: 273 Virgo Cluster galaxies and 15,294 non-Virgo galaxies. We see the local average and the 2σ confidence intervals in terms of local intrinsic scatter. The latter is shown independently in the middle panel of Figure 2.3, where we observe that the accuracy of the non-VC model drops significantly for $v \lesssim 1,500\,\mathrm{km\,s^{-1}}$ as expected from the domination of the peculiar velocities over the Hubble-flow component. Conversely, the VC model presents a slight increase in the distance with increasing radial velocity, which is possibly due to the contamination from background galaxies in the EVCC. For the same reason, the uncertainty of the inferred distances at high radial velocities for VC members is higher than that at the low radial velocities.

The convergence to the Hubble-Lemaître law is shown in the bottom panel of Figure 2.3, where we plot the ratio of the local average model ($v_{\rm vir}$; see §2.3.2.2) to the Hubble flow distance, for $v_{\rm vir} \in [200, 14,000]\ \mathrm{km\,s^{-1}}$. Two different values of H_0 are considered: $72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (HST Key Project; Freedman *et al.* 2001) and $67.8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration *et al.* 2016). We see that the z-independent distances converge to the Hubble flow distances, and agree to the local Universe estimate of the Hubble constant (HST Key Project).

Finally, we check the distance estimates in the *HECATE* against *CF3*. For 1,949 galaxies with z-dependent distances in the *HECATE*, but z-independent in *CF3*, we find a very good agreement: less than 3σ difference for 99% of the galaxies.

2.4 Multi-wavelength data and stellar population parameters

One of the main objectives for the compilation of the *HECATE* is to provide stellar population parameters for galaxies in the local Universe. To do so, we obtain photometric and spectroscopic data by cross-correlating the *HECATE* with surveys

from the infrared (IR) to the optical. In §2.4.1 we evaluate the required data to attain the most reliable galaxy properties and discuss the selection and cross-matching criteria for each survey. In §2.4.2 we describe the methodology we use for deriving the parameters from the associated multi-wavelength data.

2.4.1 Associated photometric and spectroscopic data

Star-formation rate estimates can be obtained by photometric data from IR to UV bands (or combinations of them; for a review, see Kennicutt & Evans 2012). While optical and UV-based SFR indicators are sensitive to dust absorption, IR indicators overcome this limitation by measuring the dust-reprocessed stellar emission. Although UV+IR composite SFR indicators (e.g., Hao *et al.* 2011) are now becoming more widely used (especially in the case of dwarf metal-poor galaxies) their implementation relies in the availability of integrated UV photometry. The all-sky *GALEX* UV survey does not provide integrated photometry for large, nearby galaxies, hampering the use of these SFR indicators. Therefore, we rely on mid- and far-IR indicators using *IRAS* and *WISE* photometry (see §§2.4.1.1 and 2.4.1.2), aiming at a homogeneous and as-complete-as-possible compilation of SFR estimates.

For the computation of the galaxy stellar masses, one of the most reliable photometric indicators is the K_s-band luminosity (e.g., Gardner *et al.* 1997). In order to account for the stellar-population age dependence of the mass-to-light ratio (M/L) we use calibrations that incorporate optical colours (Bell *et al.* 2003). For this reason, we obtain *2MASS* and *SDSS* photometry, as described in §§2.4.1.3 and 2.4.1.4.

Spectroscopic data can be used to estimate the metallicity of the galaxies, as well as characterise them on the basis of their nuclear activity. In §2.4.1.4 we describe the acquisition of spectroscopic data from *SDSS*.

2.4.1.1 Far-infrared: *IRAS*

We cross-link the *HECATE* galaxies to *IRAS* objects. For the cross-correlation with the *IRAS* catalogue we adopt the following approach. When a galaxy is included in the *IRAS Revised Bright Galaxy Sample* (*IRAS-RBGS*), we adopt this photometric information, which is more reliable for extended galaxies (Sanders *et al.* 2003). In total, we associate 589 galaxies with the *IRAS-RBGS* catalogue (Appendix 2.B.3). For the remaining galaxies, we use the *Revised IRAS-FSC Redshift Catalogue* (*RIFSCz*; Wang

et al. 2014) which provides a clean (excluding poor quality and cirrus sources) sample of *IRAS* galaxies at 60μm. This also gives more reliable positions than its parent *IRAS Faint Source Catalog* (*IRAS-FSC*; Moshir 1990) via the association to more recent surveys. In total, we associate 19,082 galaxies in our sample with *RIFSCz* (Appendix 2.B.4).

2.4.1.2 Mid-infrared: *WISE*

The previous cross-matches with *IRAS-RBGS* and *RIFSCz* objects incorporate *IRAS* photometry for 19,671 objects in the *HECATE* (9.6%). To obtain a more complete census of the IR emission of the galaxies in the *HECATE* sample we could also use the deeper all-sky, surveys (e.g., *WISE*, *AKARI*). However, at the time of compilation of the *HECATE*, there are no extended source catalogues of *WISE* and *AKARI* that can provide reliable flux measurements for nearby galaxies. For this reason, we use the 'forced photometry' catalogue by Lang *et al.* (2016) who extracted fluxes from *unWISE* images (Lang 2014) for *SDSS*-DR10 photometric objects using the *SDSS* apertures. We cross-correlate this catalogue with the *HECATE* by matching the *SDSS* ID, which is already specified in the *HECATE* (§2.4.1.4). As the *WISE* forced photometry catalogue is organised in *unWISE* tiles, there are galaxies in overlapping regions. For these cases, we select the data from the tile in which the galaxy is closer to the tile's centre. 123,699 *HECATE* galaxies to objects are linked to *SDSS* objects with *WISE* forced photometry. We note, however, that the use of this catalogue restricts our *WISE* photometric data to the *SDSS* footprint. *WISE* photometry is available for the wider *HECATE* sample, but as mentioned earlier it is not reliable for the resolved galaxies.

2.4.1.3 Near-infrared: *2MASS*

To incorporate *2MASS* photometry in our sample, we cross-match the *HECATE* with three *2MASS* catalogues in the following order of priority: (i) *Large Galaxy Atlas* (*2MASS-LGA*; Jarrett *et al.* 2003), (ii) *Extended Source Catalog* (*2MASS-XSC*; Skrutskie *et al.* 2006), and (iii) *Point Source Catalog* (*2MASS-PSC*; Cutri *et al.* 2012). This order ensures that for the resolved galaxies we use the most reliable measurements of their flux. Specifically, from *2MASS-LGA* and *2MASS-XSC* we obtain the *JHK* 'total' magnitudes from the extrapolated surface brightness profiles (see *2MASS-LGA*

home page* and §4.5.a.iv in the Explanatory Supplement†). From the *2MASS-PSC* we obtain the 'default' magnitudes. We note that when no uncertainty is provided, the listed magnitudes are upper limits. We link *HECATE* galaxies to 609 objects in the *2MASS-LGA*, 117,713 in the *2MASS-XSC*, and 25,224 in the *2MASS-PSC*, overall providing *2MASS* photometry for 143,546 galaxies. More details about the cross-matching procedure can be found in Appendix 2.B.5.

2.4.1.4 Optical: *SDSS*

For the cross-matching of the *HECATE* and the *SDSS*, we use the DR12 photometric catalogue, and select only primary‡ objects. We use a match radius of 3 arcsec around the *HECATE* coordinates, and we select the closest match (typical separation of the matched objects is ~0.2 arcsec), resulting in 123,711 matches.

We opt to use spectroscopic data from the *MPA-JHU DR8* catalogue (Kauffmann *et al.* 2003; Brinchmann *et al.* 2004; Tremonti *et al.* 2004), which are based on the emission-line component of the spectrum after subtracting the underlying stellar component, to estimate the metallicities and classify the galaxies on the basis of their nuclear activity. By cross-matching the catalogue with the *HECATE*, we obtain measurements for 93,714 out of the 123,711 *SDSS* objects in the *HECATE*.

The *GALEX-SDSS-WISE Legacy Catalog 2* (*GSWLC-2*) of Salim *et al.* (2016) provides SFR and $M_\star$ estimates through optical-UV spectral energy distribution (SED) fits to galaxies within the *SDSS* footprint and distance >50 Mpc. By matching 75,672 *HECATE* galaxies to *GSWLC-2* objects on the basis of their object IDs in *SDSS*, we obtain additional SFR and $M_\star$ estimates.

2.4.2 Derived parameters

The following paragraphs describe the methods employed for the estimation of parameters from the acquired multi-wavelength data (§2.4.1). An overview of the provided data is listed in Table 2.D.1.

* https://irsa.ipac.caltech.edu/applications/2MASS/LGA/intro.html
† https://old.ipac.caltech.edu/2mass/releases/allsky/doc/explsup.html
‡ www.sdss.org/dr12/help/glossary/#surveyprimary

2.4.2.1 Stellar masses

The stellar masses are estimated by combining the K_s-band luminosities of the galaxies with the appropriate mass-to-light ratios. The integrated K_s-band luminosities of the galaxies are calculated from their *2MASS* photometry and distances (we adopted 3.29 mag for the absolute magnitude of the Sun; Blanton & Roweis 2007). We exclude objects without uncertainties in their photometry, or uncertainty higher than 0.3 mag, resulting in L_K measurements for 133,017 (65%) galaxies in the *HECATE*. The K_s-band M/L ratio ($\equiv M_\star/L_{K_s}$) is computed using the calibration of Bell *et al.* (2003) which accounts for differences in the stellar populations by means of the $g-r$ colour of the galaxies:

$$\log\left(M/L\right) = -0.209 + 0.197\left(g-r\right). \tag{2.7}$$

$g-r$ colours are available for 53,171 (26%) galaxies with reliable photometry (*SDSS* flags q_mode='+' and Q=3, and uncertainties < 0.1 mag on g and r). The mean M/L ratio of the galaxies in the *HECATE* is 0.822, while the scatter is 0.091. This mean value is used for the 79,846 (39%) galaxies without *SDSS* photometry. The scatter gives us an estimation of the M/L ratio variations due to the different $g-r$ colours of the galaxies, and it can be used to assess the uncertainty on the $M_\star$ of galaxies without *SDSS* photometry. For the remainder (35%) of the *HECATE* sample that does not have K_s-band measurements in *2MASS* we do not estimate $M_\star$.

The *GSWLC-2* provides $M_\star$ derived using a different method (SED-fitting using UV to IR data; Salim *et al.* 2016). In Figure 2.4 we compare these estimates with our derived $M_\star$ using near-IR photometry. We find very good agreement (scatter of 0.21 dex), although SED-based $M_\star$ are slightly lower on average (factor of -0.11 dex), possibly due to assumptions of stellar population models, or star-formation histories.

2.4.2.2 Star-formation rates

The SFR estimates of the *HECATE* galaxies are based on measurements of IR luminosity from the *IRAS* or *WISE* surveys. These surveys provide the optimal combination of reliable, well-calibrated, SFR indicators (Kennicutt & Evans 2012), sensitivity, and sky coverage. Since the sensitivity of the two surveys varies depending on the band, we use a combination of SFR indicators depending on the availability of reliable

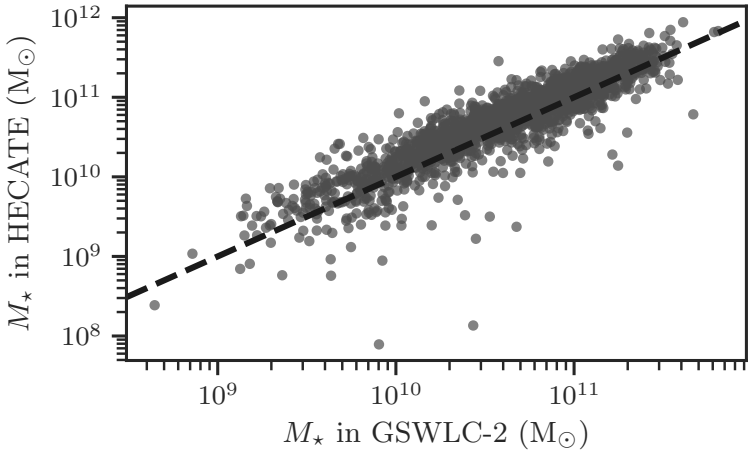


Figure 2.4: Comparison of the stellar mass estimates in the *HECATE* and the *GSWLC-2* for the common galaxies. The galaxies closely follow the 1:1 line, indicated by the black, dashed line, with an intrinsic scatter of 0.21 dex.

measurements. For *IRAS*, we use the total-IR (TIR), far-IR (FIR) and 60μm luminosities, depending on the bands with reliable *IRAS* fluxes ('FQUAL'≥2, i.e, excluding upper limits). For *WISE* we use the monochromatic Band-3 (W3; 12μm) and Band-4 (W4; 22μm) fluxes from the *WISE* forced photometry catalogue as discussed in §2.4.1.2. SFR estimates are not provided for objects with uncertainties greater than 0.3 mag, or those that were considered as point sources in the analysis of Lang *et al.* (2016)*.

For completeness, we provide in our catalogue SFR measurements based on all indicators (Table 2.2) available for each galaxy (including the SED-based SFR from the *GSWLC-2*). This is particularly important since the various surveys used to derive the SFRs, cover different subsets of the *HECATE*. In order to have consistent SFRs for the largest possible set of objects, and given the fact that different indicators often result in systematic offsets in the derived SFRs, we also provide a homogenised SFR (SFR_{HEC}), calculated as follows.

First, we account for offsets between the different SFR indicators (Table 2.2) by calculating their ratio with respect to the TIR-based SFR which we consider as a reference. The mean ratio for each indicator is adopted as the correction factor. Figure 2.5 shows comparisons between the SFR indicators, also giving the scaling factor, standard deviation, and the number of galaxies used in each comparison. The

* Using the galaxies with *WISE* and *IRAS* photometry we found that the *WISE* SFRs are significantly lower for sources with the flag 'treated_as_pointsource' set in the catalogue of Lang *et al.* (2016).

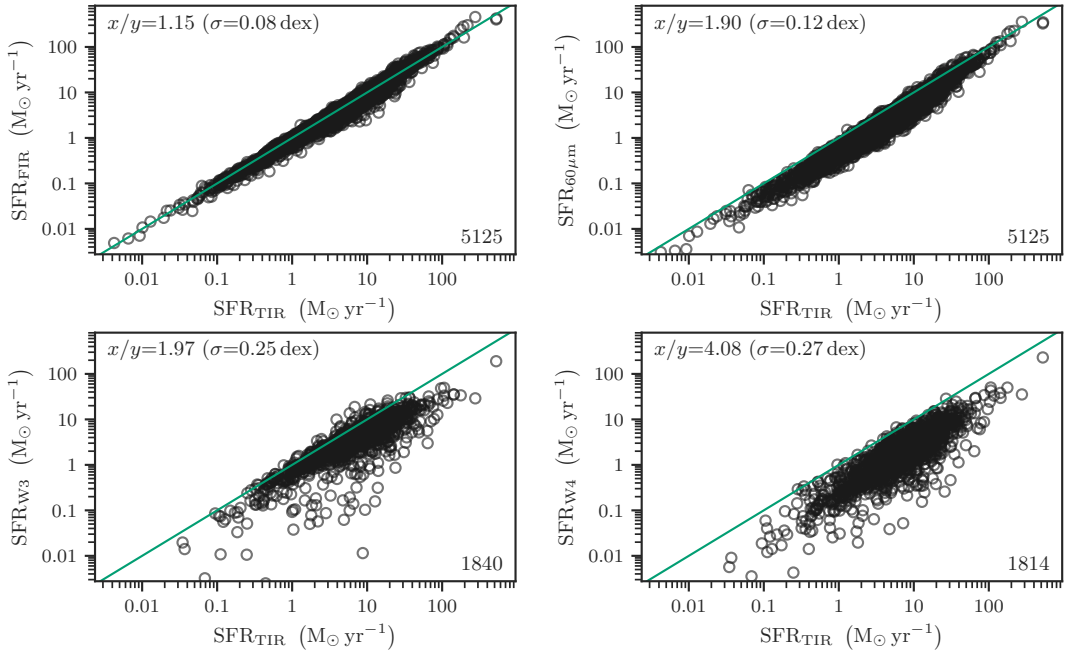


Figure 2.5: Comparison between the TIR-SFR indicator and the other four indicators used in the *HECATE* (not rescaled as in the computation of SFR_{HEC}); the 1:1 line is shown as a green line. For each SFR indicator, the linear scaling factor and the scatter are reported in the top left corner, while the number of overlapping galaxies used for the scaling is reported in the bottom right corner. The four SFR indicators scale with the reference indicator (TIR) well, and present intrinsic scatter upto 0.27 dex (typical of photometric SFR estimates; Kennicutt & Evans 2012). The scaling factors are used for the computation of the homogenised SFR column in the *HECATE*.

homogenised SFR of an object in the *HECATE* is the TIR-SFR if available, otherwise we use, in order of preference, the rescaled FIR-SFR, 60μm, 12μm and 22μm based SFR. Although the 22μm-band (probing hot dust associated with young star-forming regions) is a better-calibrated SFR than the 12μm-based one (probing emission from polycyclic aromatic hydrocarbons; e.g., Parkash *et al.* 2018), preference is given to the latter due to the higher quality of the W3 *WISE* data (e.g., Cluver *et al.* 2017). We note that no rescaling is performed in the individual SFR indicator columns.

The TIR luminosity includes emission in the 100μm band, which in the case of galaxies with low specific SFR may have a non-negligible contribution from stochastically heated dust from older stellar populations (e.g., Galliano *et al.* 2018). Although this may overestimate the SFR in early-type galaxies, it is a widely used and well understood SFR indicator that gives reliable SFR for actively star-forming galaxies,

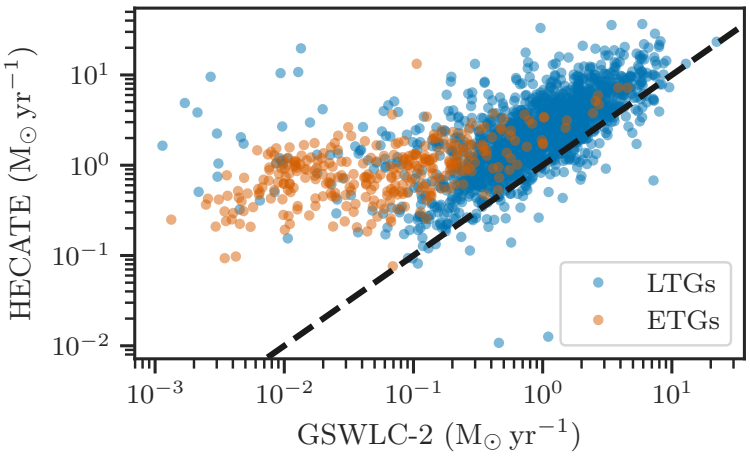


Figure 2.6: Comparison between the homogenised SFRs and the SED-based SFRs from the *GSWLC-2* for early (orange) and late-type (blue) galaxies. In the bulk of the sample, mainly consisting of late-type galaxies, the *GSWLC-2* underestimates the SFR due to the lack of the dust component associated with the star-forming activity, whereas in early-type galaxies, the *HECATE* overestimates the SFR due to the dust emission caused by the stochastic heating from old stellar populations, rather than star-formation.

which are the majority of the *HECATE* (fig. 2 in Kovlakas *et al.* 2020). We note that the catalogue provides all SFR indicators (Appendix 2.D) before rescaling, and the homogenised SFR, where indicators were rescaled according to the procedure described above. A flag is provided, denoting which indicator was used in the homogenised SFR (cf. Table 2.2). Therefore, based on the scaling factors reported in the table, one can translate the provided SFR to the reference indicator of their choice.

In Figure 2.6 we compare the homogenised SFRs against the SED-based SFRs from the *GSWLC-2*. We see that at SFR$\gtrsim 0.1$ $M_\odot$ yr^{-1} (typical for star-forming galaxies), the *HECATE* provides SFRs that scale with, but are a factor of $\simeq$2.3 larger than those of *GSWLC-2*. This could be because the SEDs used in the *GSWLC-2* do not include IR emission above 22μm, therefore missing the dominant, relatively cold, dust component associated with star-forming activity (probed in the $\sim$60μm band). As discussed above, the IR-based SFRs may overestimate the SFR in low specific-SFR galaxies, which can explain the flattening observed at low SFRs, and the difference between early-type and late-type galaxies.

Table 2.2: The five different SFR indicators used in the *HECATE*.

Survey / bands (1)	N_{gal} (2)	Calibrations (3)	Scaling in SFR_{HEC} (4)	Flag (5)
IRAS 25, 60, 100μm	5,125	Dale & Helou 2002; Kennicutt & Evans 2012	(reference)	RT/FT*; 5,125
IRAS 60, 100μm	5,721	Helou & Walker 1988; Kennicutt 1998	1.15 (0.08 dex)	FF; 596
IRAS 60μm	19,671	Rowan-Robinson 1999	1.90 (0.12 dex)	R3/F3*; 13,950
WISE 12μm	81,948	Cluver *et al.* 2017	1.97 (0.25 dex)	W3; 72,726
WISE 22μm	46,078	Cluver *et al.* 2017	4.08 (0.27 dex)	W4; 1,872

Description of columns: (1) the survey (*IRAS* or *WISE*), and the photometric bands used for the computation of the flux; (2) the number of galaxies for which the SFR indicator is computed; (3) references to the definition of the composite band and SFR scaling calibration; (4) The scaling factor used only for the 'homogenisation' of the SFR indicator with respect to the TIR indicator. In parenthesis we give the scatter of the 'homogenisation' relation; (5) the flag (in the column 'logSFR_flag'; see Table 2.D.1) and the number of galaxies used to calculate the homogenised SFR. **Note**: * the first letter indicates whether the photometry was taken from the *IRAS-RBGS* (R) or *RIFSCz* (F).

2.4.2.3 Metallicity estimates

To measure the gas-phase metallicities for our sample, we use the optical emission-line fluxes provided in the *MPA-JHU DR8* value-added 'galSpecLine' catalogue (see details for methods in Brinchmann *et al.* 2004) based on the *SDSS*-DR8 data. Especially relevant for measuring accurate nebular emission lines, this catalogue applies stellar-population synthesis models to accurately fit and subtract the stellar continuum, including stellar absorption features. We calculate the gas-phase metallicities, $12 + \log(O/H)$, using the Pettini & Pagel (2004) O3N2 (henceforth, PP04 O3N2) prescription, which has been shown by Kewley & Ellison (2008) to be robust (i.e. it can trace a wide range of metallicities, it has relatively low scatter, and most importantly it is less sensitive to extinction effects than other indicators). Our metallicity analysis is subject to the quality of the [O III], [N II], Hβ, or Hα emission lines and the PP04 O3N2 relation limitations. Therefore, we set the following flags (see column 'flag_metal' in Appendix 2.D) to mark uncertain results: sources with '1' have O3N2 >2 ratios (670 sources), where the PP04 O3N2 relationship is invalid, and therefore the extrapolated metallicities are highly uncertain; '2' marks emission lines with low signal-to-noise (σ<3; 882 sources). Therefore, only sources with flags set to '0' have reliable metallicity measurements (62,728 sources). Objects without metallicity estimates are flagged with

'-1' (140,453 sources).

2.4.2.4 Nuclear activity

Using the *SDSS*-DR8 emission-line data (see §2.4.1.4 and §2.4.2.3), we identify AGN based on the location of the galaxies in the emission-line ratio diagnostic of Stampoulis *et al.* (2019). This diagnostic takes into account all available line ratios in order to provide a single robust activity classification that avoids the contradictory classifications that can be obtained from the use of the traditional two-dimensional line-ratio diagrams. We consider the ([S II $\lambda\lambda$6717, 6731Å]/Hα, [N II λ6584Å]/Hα, [O III λ5007Å]/Hβ) three-dimensional diagram and when we have reliable measurements for the [O I] λ6300Å line we use the four-dimensional ([S II]/Hα, [N II]/Hα, [O I]/Hα, [O III]/Hβ) diagram. In this way, we provide nuclear activity classification for 64,280 (31%) galaxies with signal-to-noise ratio greater than 2 in the emission lines used. Out of these 64,280 galaxies, 9,987 (15%) are characterised as AGN, leaving a non-AGN sample of 54,293 galaxies.

One of the motivations for the compilation of this galaxy catalogue, was the study of X-ray source populations in 'normal' (i.e., non AGN-hosting) nearby galaxies. Therefore, we also include the AGN classifications from She *et al.* (2017) who studied galaxies that had been observed with *Chandra* at distances less then 50 Mpc. In total, we obtain classifications for 716 galaxies.

Finally, we combine the classifications in a single estimate. For galaxies with classifications from only one of the two sources, we adopt them as they are. For galaxies both in the *SDSS* and She *et al.* (2017) sample, they are characterised as AGN if they are classified as such by either of the two sources, otherwise as non-AGN. The *SDSS*, She *et al.* (2017), and the combined classifications are all provided in the catalogue (for 64,280, 716, and 64,910 objects respectively, leaving 139,823 galaxies without classification).

2.5 Discussion

2.5.1 Comparison with other catalogues

Out of the available all-sky galaxy catalogues only the *Galaxy List for the Advanced Detector Era* (*GLADE*; Dálya *et al.* 2018), the *Mangrove* (Ducoin *et al.* 2020), and the

Census of the Local Universe (*CLU*; Gehrels *et al.* 2016; Cook *et al.* 2019) are similar in scope (i.e., offer multi-wavelength photometry and galaxy characterisation) as the *HECATE*.

The *GLADE* galaxy catalogue provides coordinates, distances, and photometry in the B, J, H, and K-bands by cross-matching five catalogues: *HyperLEDA*, *2MASS*-XSC, *GWGC*, the *2MASS photometric redshift catalogue*, and *SDSS*-DR12Q. Without an explicit limit on z, it is an ideal tool for low-redshift cosmology, and studies of distant transient events such as long GRBs. A recent extension of the *GLADE* is the *Mangrove* catalogue, which provides $M_\star$ estimates via mid-IR photometry obtained by cross-matching the *GLADE* and the *AllWISE* catalogue.

Over the distance range covered by the *HECATE*, the completeness of *GLADE* and *Mangrove* in terms of the B-band luminosity is similar to that of the *HECATE* (cf., fig. 2 in Dálya *et al.* 2018 and Figure 2.8). However, the two catalogues do not include size information for the galaxies, limiting their usability for the association of host galaxies with sources from serendipitous and all-sky surveys (e.g., Webb *et al.* 2020). Another important difference between the *HECATE* and the *GLADE* or *Mangrove* is that the *HECATE* provides robust distances, SFRs based on a wide suite of indicators, as well as, homogenised SFRs that bridge the systematic differences between the individual indicators, and integrated *2MASS* and *WISE* photometry for nearby galaxies.

The *CLU* catalogue has been progressively constructed since 2016 to aid the identification of GW hosts (Gehrels *et al.* 2016), and provide a census of emission-line galaxies with D<200 Mpc using new observations (Cook *et al.* 2019). Including information from the *NED*, *HyperLEDA*, Extragalactic Distance Database, *SDSS*-DR12, 2dF Galaxy Redshift Survey, the Arecibo Legacy Fast ALFA, *GALEX*, and *WISE*, it provides multi-wavelength data, SFRs and $M_\star$ based on *WISE* photometry. However, for studies of nearby galaxies, the *CLU* has the same limitations as in the case of *GLADE*: it does not provide size information on the sample galaxies, and the *WISE*-based photometry is problematic for nearby, extended objects (§2.4.1.2).

Concluding, the *HECATE* provides robust distances (an important parameter for nearby galaxies; see §2.3), and additional data that are not readily-available in the other catalogues: reliable homogenised SFRs, metallicities, as well as morphological and AGN classifications.

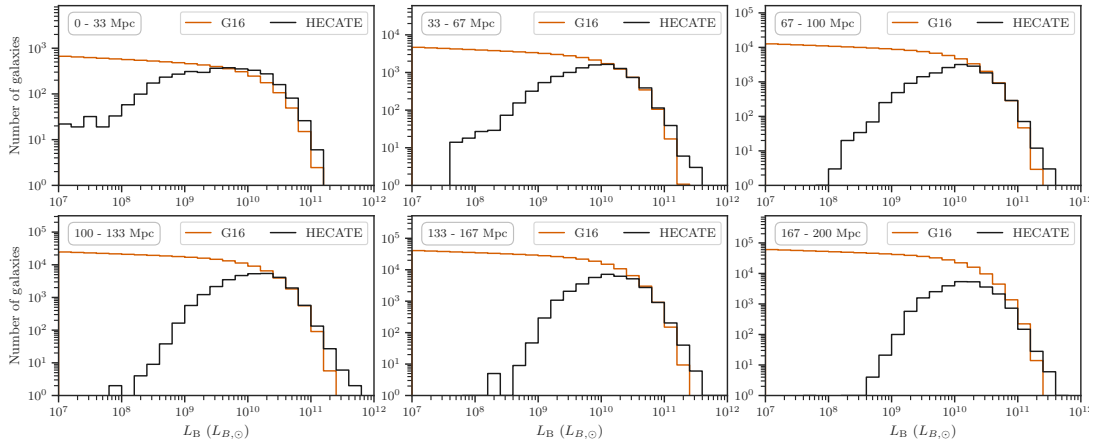


Figure 2.7: The distribution of the *B*-band luminosities of the galaxies in the *HECATE* at different distance ranges (top left boxes), and comparison against the expectation from the L_B LF in Gehrels *et al.* (2016).

2.5.2 Completeness

The completeness of the *HECATE* cannot be robustly calculated due to the unknown selection function of the *HyperLEDA*, which is further complicated by the selection effects introduced by the other catalogues that it is cross-correlated with. However, we can obtain an estimate of the completeness by comparing the distribution of *B*-band luminosities with the expectation from the galaxies LF, following the approach of Gehrels *et al.* (2016) and Dálya *et al.* (2018). Using the same Schechter LF as in the aforementioned papers*, we compute the expected number of galaxies in different bins of luminosities and distances, shown in orange in Figure 2.7, which we compare with the number of galaxies in the *HECATE* in the respective bins (black). We find that the *HECATE* is complete down to $L_B \sim 10^{9.5} L_{B,\odot}$ at distances less than 33 Mpc, and down to $L_B \sim 10^{10} L_{B,\odot}$ at $67 < D < 100$ Mpc. However, at distances greater than 167 Mpc the *HECATE* suffers by incompleteness even at the high-end of the LF.

Since many applications of the *HECATE* are related to the stellar content of the galaxies, we can quantify its completeness in terms of the ratio of the integrated *B*-band luminosity of galaxies at distance *D*, with respect to the mean L_B-density of the local Universe. This approach has been followed in several studies of nearby samples of galaxies: Kopparapu *et al.* (2008); White *et al.* (2011); Gehrels *et al.* (2016); Dálya

* $\Phi = 1.6 \times 10^{-2}\ h^3\ \mathrm{Mpc}^{-3}$, $a = -1.07$, and $L_\star = 1.2 \times 10^{10}\ \mathrm{L}_{B,\odot}$ (cf. Gehrels *et al.* 2016). We adopt $h = 0.7$ as an intermediate value between the published H_0 calibrations.

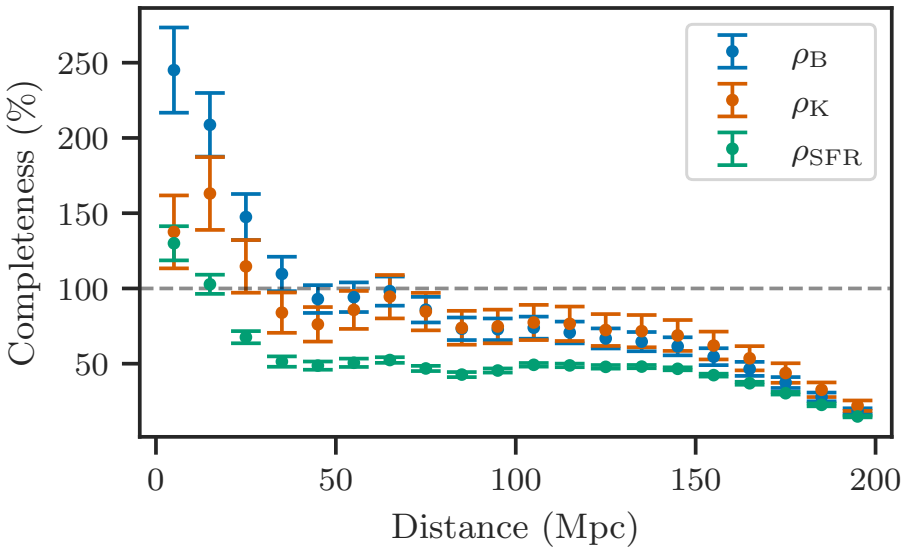


Figure 2.8: The completeness of the *HECATE* in terms of the included *B*-band, K_s-band, and SFR density with respect to the expectation from observational estimates. The completeness is between 50% and 100% within 150 Mpc. At small distances, the completeness exceeds 100 per cent because of the overdensity in the neighbourhood of the Milky Way (cf. Gehrels *et al.* 2016).

et al. (2018). We adopt the mean L_B density of $(1.98 \pm 0.16) \times 10^8$ Mpc^{-3} (Kopparapu *et al.* 2008) that was used by the aforementioned works. To account for the different sources of uncertainties. we sample from the distributions of the various quantities involved in the computation (i.e., the mean L_B-density, and the galaxy distances), and compute the completeness in bins of 10 Mpc. This is performed for 10,000 iterations to obtain the mean and standard deviation of the completeness as a function of the distance. The L_B-completeness is shown by blue error bars in Figure 2.8. We find that the *HECATE* is >75% complete in terms of the blue light at D<100 Mpc, and ~50% at D~170 Mpc. The completeness above 100% at small distances (D<30 Mpc) is the result of the over-density in the neighbourhood of the Milky Way.

Similarly, we calculate the completeness of the *HECATE* in terms of the $M_\star$. For this reason we perform the same exercise with the K_s-band luminosity, which is a tracer of the $M_\star$ of the galaxies. We adopt a K_s-band luminosity density of $5.8 \times 10^8 h L_{K,\odot}$ Mpc^3 (Bell *et al.* 2003). The result is similar to the L_B-completeness as shown by orange in Figure 2.8, exhibiting both the excess at small distances and the cut-off at large distances.

The completeness in terms of the SFR is calculated in the same way (shown by green points in Figure 2.8), adopting a local Universe SFR density of 0.015 $M_\odot\ yr^{-1}\ Mpc^{-3}$ (Madau & Dickinson 2014), and using the homogenised SFR for the *HECATE* galaxies. In this case, the *HECATE* is incomplete at all distances in its regime, with ~50% completeness at $30<D<150$ Mpc. The higher incompleteness in SFR with respect to the other parameters (L_B and $M_\star$) stems from the fact that the *WISE*-based SFRs in the *HECATE* do not have all-sky coverage since they are based on forced photometry on *SDSS* objects. Nevertheless, due to the all-sky coverage of *IRAS* and despite its shallowness, it covers more than 50% of the star-forming activity in the Galactic neighbourhood.

2.5.3 Limitations

The parent sample of the *HECATE*, the *HyperLEDA* database, includes objects and related data, from hundreds of surveys with different sky coverage and sensitivity limits. Therefore, the selection function of the *HyperLEDA*, and as a consequence, that of the *HECATE*, is intractable (§2.5.2). Generalisations based on the provided galaxy compilation should be treated carefully.

At low distances ($D\lesssim20$ Mpc) peculiar velocities dominate the Hubble flow (§2.3.2). This is accounted for by the regression model for estimating distances based on the recession velocities of the galaxies, however the increased scatter reduces the accuracy of the inferred distances for velocities $v_{vir}\lesssim 1,500\,km\,s^{-1}$ (Figure 2.3). This can be remedied by measuring z-independent distances for the nearby galaxies. In addition, there are a few cases where distance measurements are significantly different from the Hubble-flow distance*. The causes of these discrepancies are diverse and difficult to identify in most cases (e.g., problematic distances due to biases in distance indicators, wrong redshifts because of superimposed stars, typos, etc.) In the future, the methods for estimation of distances will include special treatment for such outliers.

Furthermore, the derived stellar population parameters are based on multi-wavelength data from combinations of surveys and calibrations. The statistical treatments presented in this paper (e.g., homogenisation of SFR estimates, fixed M/L ratio for galaxies without M/L estimate) provide estimates of stellar populations for a large

* e.g., NGC 5434 is reported to have a distance of 3.8 Mpc both in *NED-D*, and subsequently in the *HECATE*, but its z implies $D\approx70$ Mpc.

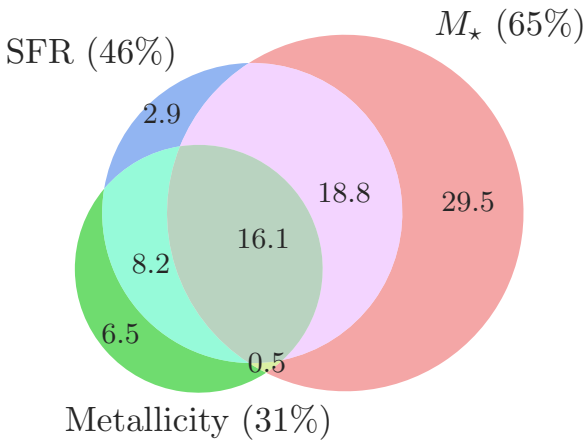


Figure 2.9: Venn diagram of the coverage of the stellar population parameters, SFR, $M_\star$, and metallicity, in the *HECATE*. The per cent coverage for each is reported next to its label, while the numbers in the coloured areas denote the percentage for the different combinations of parameters.

fraction of galaxies in the local Universe. While this allows for statistical studies of large galaxy samples, or quick searches for objects of interest, more accurate methods ought to be preferred when focusing on individual galaxies.

The IR-based SFR estimates are based on calibrations that assume 'normal' star-forming galaxies. In the case of quenched, early-type galaxies, the SFRs may be overestimated (e.g., Hayward *et al.* 2014). Indicators based on optical-UV SED analysis could be more reliable for these galaxies.

One of the most important limitations of the *HECATE* is its non-uniform coverage in terms of the SFR and $M_\star$. In Figure 2.9 we show the coverage of stellar population parameters in the *HECATE*. SFR, $M_\star$, and metallicity estimates are available for 46%, 65%, and 31% of the galaxies, respectively. Currently, the *WISE* photometry is obtained through the forced photometry catalogue of Lang *et al.* (2016) which is limited to the *SDSS* footprint. While this is driven by the need for accurate photometry for the extended galaxies (which are the majority of the *HECATE* galaxies), it leaves a significant fraction of the sample without sensitive IR photometry that could provide reliable and uniform SFR measurements. For specific cases, this limitation can be remedied by including in the analysis data from additional catalogues. In a future version of the *HECATE* we will apply the forced photometry method to all galaxies in the *HECATE*, thus providing robust stellar population parameters. In addition,

incorporation of additional photometry and spectroscopy from other surveys (e.g., Pan-STARSS, LAMOST; Chambers *et al.* 2016; Deng *et al.* 2012), will increase the multi-wavelength, AGN classification, and metallicity coverage of the *HECATE*. This will also allow the computation of SED fits, that will provide additional SFR and $M_\star$ estimates for the galaxies.

2.5.4 Applications

The motivation for creating an all-sky galaxy catalogue with positions, sizes, multi-wavelength data, and derived parameters (e.g., SFR, $M_\star$, metallicity) was to enable several applications relying on the initial characterisation of sources in the context of the host galaxy, or identifying counterparts of transient events for follow-up observations. In this section we outline some specific use cases.

2.5.4.1 Application to all-sky and serendipitous surveys

The distance limit of the *HECATE* and the large array of the information it provides make it an ideal sample for designing wide-area multi-wavelength surveys, or characterising sources within. For example, it can form the baseline sample for realistic simulations of the data expected to be obtained with future surveys (see Basu-Zych *et al.* 2020, for an application to the *eROSITA* survey), but also it can be a reference sample for the initial characterisation of newly-identified sources (e.g., with Dark Energy Survey: Flaugher 2005; Large Synoptic Survey Telescope, LSST survey: Ivezić *et al.* 2019).

A demonstration of the potential of the *HECATE* is given in Kovlakas *et al.* (2020), a comprehensive study of ultraluminous X-ray sources in the local Universe based on the *Chandra Source Catalog 2.0* (Evans *et al.* 2010). The positional and size information available in *HECATE* allowed the association of X-ray sources with their host galaxies and the robust estimation of the fraction of interlopers. In addition the SFR, $M_\star$, and metallicity information was used to derive scaling relations between the ULXs and the stellar populations in their host galaxies. The special treatment of nearby galaxies (e.g., extended photometry) in the *HECATE* was essential for the science in this project since the target sample was limited in a volume out to 40 Mpc. Similarly, the combination of *HECATE* with *XMM-Newton* has been is the basis for the largest study of the X-ray scaling relations of galaxies (Anastasopoulou et al., in preparation),

and the largest *XMM-Newton* census of ULXs in nearby galaxies up to date (Bernadich et al., in preparation).

2.5.4.2 Application in search of EM counterparts to GW sources

All-sky galaxy catalogues are crucial for the timely identification of electromagnetic (EM) counterparts to GW sources (e.g., Nissanke *et al.* 2013; Gehrels *et al.* 2016). This is a key step for constraining their nature (e.g., Abbott *et al.* 2017b), understanding the formation and evolution of their progenitors (e.g., Kalogera *et al.* 2007; Abbott *et al.* 2017c), or even using them as standard 'sirens' to measure the Hubble constant (e.g., Schutz 1986; Chen *et al.* 2018; Abbott *et al.* 2017a).

The poor localisation of GW sources by the contemporary GW detectors ($\gtrsim$100 deg^2; Abbott *et al.* 2020), makes the search for EM counterparts a daunting task. The adopted solution is to perform targeted follow-up observations of a list of potential hosts prioritised based on properties such as their distance, or the parameters of their stellar populations (e.g., Kanner *et al.* 2008; Nuttall & Sutton 2010; Gehrels *et al.* 2016; Kasliwal *et al.* 2017; Arcavi *et al.* 2017; Del Pozzo *et al.* 2018; Yang *et al.* 2019; Salmon *et al.* 2020; Wyatt *et al.* 2020).

This approach has already led to the compilation of galaxy catalogues that provide in addition to positions and distances, photometric information (as proxies to SFR; e.g., Kopparapu *et al.* 2008; White *et al.* 2011; Gehrels *et al.* 2016), or directly SFR and $M_\star$ determinations (e.g., Dálya *et al.* 2018; Cook *et al.* 2019; Ducoin *et al.* 2020). This is driven by models which predict that GW populations scale with SFR (e.g., Phinney 1991), and/or $M_\star$ (e.g., Mapelli *et al.* 2017; Artale *et al.* 2019; Toffano *et al.* 2019; Adhikari *et al.* 2020). However, these catalogues lack information on metallicity which can be an important factor in the GW rates (e.g., O'Shaughnessy *et al.* 2017; Mapelli *et al.* 2018; Artale *et al.* 2019; Neijssel *et al.* 2019; Artale *et al.* 2020b; Bavera *et al.* 2020).

The *HECATE*, having a distance limit (~200 Mpc) that is sufficient for searches of EM counterparts to GW sources from binary neutron stars (BNS) until the mid-2020s (e.g., Buikema *et al.* 2020), and providing stellar population parameters, can be used for assigning likelihoods to putative GW hosts for observational follow-up campaigns. In this section, we use as an example the GW event GW170817, the only case of verified EM counterpart of a BNS, to

1. illustrate the use of the *HECATE* in producing priority lists of galaxies for EM counterpart searches,
2. study the effect of the different pieces of information (direction, distance, and stellar population parameters) in the prioritisation of host candidate galaxies,
3. assess, post facto, the ability of various schemes in giving high priority to the host galaxy of GW170817, NGC 4993.

The priority lists are the result of ordering the galaxies based on their probability of being the hosts,

$$P \propto P_{\rm 3D} \times G_{\rm intr} \tag{2.8}$$

where $P_{\rm 3D}$ is the volume-weighted probability given the position and distance of the galaxy, and $G_{\rm intr}$ is a factor (or grade) which scales with the probability for a galaxy to host a GW event given its intrinsic properties (e.g., $M_\star$ or SFR proxy, or merger rate).

As a first step, we acquire the HEALPix map (Górski *et al.* 2005) produced by BAYESTAR (Singer *et al.* 2016) which contains the 2-D localization probability, i.e. the probability that the GW event is on a specific direction of the sky, and the corresponding distance probability distribution. By cross-matching the *HECATE* with the HEALPix map we find 2,249 candidate host galaxies in the 99.9% region of GW170817. As the 'directional', namely the 2-D probability of the galaxy, $P_{\rm 2D}$, we assign the value of the HEALPix pixel which contains the centre of the galaxy.

The 3-D probability, $P_{\rm 3D}$=$P_{\rm 2D}$×$P_{\rm d}$, is computed by combining the $P_{\rm 2D}$ with the GW event distance probability density ($P_{\rm d}$) for the corresponding pixel in the HEALPix map, and the distance of the galaxy in the *HECATE*.

Subsequently, the 3-D probabilities are multiplied by 'astrophysical' terms ($G_{\rm intr}$) which are assumed to be proportional to the merger rate of BNSs, and therefore the probability of a merger. The astrophysical terms are generally parametrised in terms of the $L_{\rm B}$ (cf. Arcavi *et al.* 2017; Salmon *et al.* 2020), stellar mass (cf. Ducoin *et al.* 2020), and the theoretical predictions on the merger rate of BNS, as a function of different combinations of the stellar population parameters: (i) $n(M_\star)$, (ii) $n(M_\star, {\rm SFR})$, and (iii) $n(M_\star, {\rm SFR}, Z)$, based on the results of Artale *et al.* (2020b) for $z = 0$ (cf. their table 1), where Z is the metallicity of the galaxy.

Since the three stellar population parameters may not be known for all galaxies in the HEALPix map, we also use a 'combined' estimate, where the appropriate merger rate is used depending on the available information:

$$n_{\rm comb} = \begin{cases} n(M_\star, {\rm SFR}, Z) & \text{if } M_\star, \text{ SFR and } Z \text{ are defined} \\ n(M_\star, {\rm SFR}) & \text{if } M_\star \text{ and SFR are defined} \\ n(M_\star) & \text{if } M_\star \text{ is defined} \end{cases} . \tag{2.9}$$

Finally, in order to include in our analysis galaxies without $M_\star$ estimates (for which n cannot be inferred), we employ the weighting scheme of Ducoin *et al.* (2020):

$$P_{\rm Du} \propto P_{\rm 3D}\,(1 + \alpha n_{\rm comb})\,, \quad \text{where} \quad \alpha = \frac{\sum P_{\rm 3D}}{\sum P_{\rm 3D} n_{\rm comb}}. \tag{2.10}$$

The quantities, $P_{\rm 2D}$, $P_{\rm 3D}$, $P_{\rm 3D}{\times}L_{\rm B}$, $P_{\rm 3D}{\times}M_\star$, $P_{\rm 3D}{\times}n(M_\star)$, $P_{\rm 3D}{\times}n(M_\star, {\rm SFR})$, $P_{\rm 3D}{\times}n_{\rm comb}$, and $P_{\rm Du}$, are used to produce priority lists of the host galaxy candidates, to test the aforementioned schemes for prioritising candidate host galaxies (Table 2.3). The scheme which accounts for the metallicity dependence of the merger rate is omitted due to lack of metallicity estimates in the sky region of the GW event.

We find that NGC 4993 is given the highest priority by the schemes involving the $L_{\rm B}$ or $M_\star$, and second priority for those also involving the SFRs*. Except for the priority lists based only on the 2-D or 3-D position, the lists feature the same top-five galaxies as in the first prioritisation list published after the GW170817 alert (Kasliwal *et al.* 2017) based on the *CLU* catalogue (Gehrels *et al.* 2016). The same holds for the top-three galaxies reported in (i) Artale *et al.* (2020a) who use $M_\star$ and SFR estimates from the *Mangrove* catalogue (Ducoin *et al.* 2020), and (ii) in Yang *et al.* (2019) who used the B-band luminosity from *GLADE* (Dálya *et al.* 2018) as the 'astrophysical' term. The top-ranked *HECATE* galaxies based on the different prioritisation schemes, agree to a high degree with the results of the same schemes in Ducoin *et al.* (2020) using the *Mangrove* catalogue.

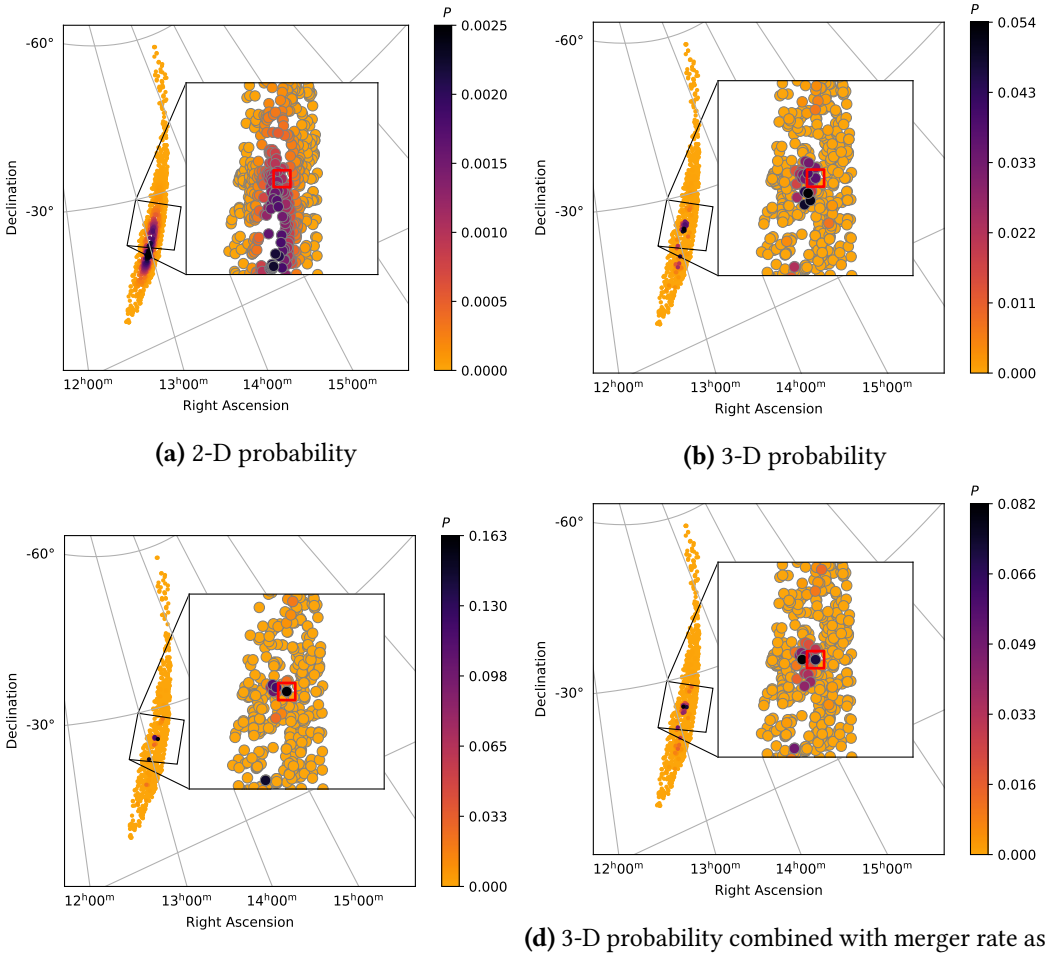
While in the case of GW170817 there is no significant difference between the use of B-band luminosity, $M_\star$ or the fits with $M_\star$ and SFR (NGC 4993 was always first, or

* Following §2.5.3, because GW170817 falls outside the *SDSS* footprint, for this application we supplement the *IRAS* photometry with mid-IR photometry from the *AllWISE* catalogue.

Table 2.3: Prioritisation lists of host galaxy candidates (first five sources), and computed probabilities based on different schemes. The true host (bold text), NGC 4993, is successfully recovered as first or second most probable host galaxy once the astrophysical information is accounted for.

Galaxy[(i)]	P_{2D}	Galaxy[(ii)]	P_{3D}
PGC4690279	0.002	ESO508-004	0.054
PGC3799401	0.002	ESO575-055	0.051
PGC3798804	0.002	ESO575-053	0.049
PGC4690296	0.002	PGC4692149	0.045
PGC4690280	0.002	PGC169673	0.045
Galaxy	with L_B	Galaxy	with $M_\star$
NGC4993	0.096	**NGC4993**	0.163
ESO508-019	0.079	NGC4830	0.148
IC4197	0.074	IC4197	0.119
NGC4830	0.073	NGC4970	0.115
NGC4970	0.072	NGC4968	0.103
Galaxy	with $n(M_\star)$	Galaxy	with $n(M_\star, \mathrm{SFR})$
NGC4993	0.164	NGC4968	0.180
NGC4830	0.151	**NGC4993**	0.135
IC4197	0.121	NGC4830	0.102
NGC4970	0.117	IC4187	0.100
NGC4968	0.102	NGC4970	0.087
Galaxy	with n_{comb}	Galaxy	P_{Du}
NGC4968	0.180	NGC4968	0.082
NGC4993	0.135	**NGC4993**	0.071
NGC4830	0.102	NGC4830	0.050
IC4197	0.100	IC4197	0.049
NGC4970	0.087	NGC4970	0.044

Notes: [(i)] The rank of NGC 4993 is 461. [(ii)] The rank of NGC 4993 is 7.



(a) 2-D probability

(b) 3-D probability

(c) 3-D probability combined with $M_\star$

(d) 3-D probability combined with merger rate as a function of $M_\star$ and SFR, using the Ducoin *et al.* (2020) weighting scheme.

Figure 2.10: (a) The sky distribution of the 2,248 galaxies in the *HECATE* that lie inside the 99.9% CR of the 2-D localisation map for the GW170817 event. The inset zooms in the region of the galaxy NGC 4993 (red square). The colour indicates the normalised 2-D probability across all the host candidates. In **(b)**, the 3-D probability is used accounting for the distance estimates at each direction in the sky. In **(c)**, the 3-D probability is multiplied by the stellar mass. Similarly, in **(d)** the 3-D probability is multiplied by the merger rate as computed by the fits in Artale *et al.* (2020b), and the weighting scheme of Ducoin *et al.* (2020) is used to allow galaxies without stellar population parameters estimates to enter in the prioritisation list. Note that in (a) and (b) the contrast of the highest-probability galaxies to the rest is small. The introduction of the 'astrophysical' terms in (c) and (d), gives prominence to NGC 4993.

second with a small difference in the probability), refined prioritisation schemes will be important for the quick identification of EM counterparts of future BNS coalescence signals with poorer localisation or at higher distances. The *HECATE*, making readily available a large set of intrinsic properties for the candidate host galaxies, offers versatility in the choice, design, and assessment of different priority schemes.

We note that the practice of initially narrowing down the galaxy sample by deciding on a confidence region based on the 2-D probability, increases the risk of missing the true host. This is indicated by the rank of NGC 4993 in the P_{2D}-based priority list (461), and the fact that total 2-D probability of galaxies closer to the centroid (a few degrees from NGC 4993) is as high as 75%. We suggest using the full galaxy catalogue together with priority schemes involving distance (and astrophysical information where possible). For example, the inclusion of the distance information in P_{3D}, promotes NGC 4993 to the 7th position, and shifts the centroid by a few degrees (see panel (b) of Figure 2.10), a consequence of the non-homogeneity of the Universe at the distance of the event (out of the 15 galaxies in Table 2.3 for schemes with 3-D positional term, 10 are considered members of the NGC4970 group; Kourkchi & Tully 2017).

2.5.4.3 Application in short gamma-ray bursts

Another manifestation of BNS mergers are short GRBs (sGRBs; e.g. Tanvir *et al.* 2013) as it has been shown by the association of GW170817 to GRB170817A (Goldstein *et al.* 2017). The identification of the host galaxies of sGRBs is important for two reasons: (a) connecting their populations with the star-formation history of their host galaxies, and (b) measuring the displacement of the GRBs from their host galaxies. The former is key for modelling the evolutionary paths of sGRBs and their cosmological evolution (e.g., Leibler & Berger 2010; Selsing *et al.* 2018; Abbott *et al.* 2017c). The latter is important for constraining the effect of kicks in the populations of sGRBs (e.g., Zevin *et al.* 2019). and studying the enrichment of the interstellar medium in r-process elements (e.g., Andrews & Zezas 2019). The *HECATE* can provide the initial information required to quickly associate a GRB with their host galaxy which is also important for prompt follow-up observations.

2.5.4.4 Localisation of neutrino and cosmic-ray sources

In the case of neutrino events, the large error-box of their localisation poses the same challenges as the GW detections (e.g., Krauß *et al.* 2020). Therefore, a catalogue of galaxies, which is as complete as possible, can be a valuable resource for the identification of their origin when one could follow a similar approach as the prioritised host list developed for GW events (§2.5.4.2). Furthermore, the availability of multi-wavelength data, stellar population parameters, and nuclear activity classifications is particularly useful for the identification of potential neutrino sources (see IceCube Collaboration 2018, and references therein).

The same holds for the case of cosmic-ray detections, which also have very large error-circles (e.g., Pierre Auger Collaboration 2015). Recent studies of anisotropy in the arrival direction of high-energy cosmic rays, and the existence of a dipole at high Galactic latitude, indicate that their origin is neither exclusively Galactic, nor cosmological (e.g., Pierre Auger Collaboration 2017). Since the Greisen-Zatsepin-Kuzmin effect limits the propagation of high-energy cosmic rays to $\lesssim$100 Mpc (e.g., Bhattacharjee 2000), the *HECATE*, as an all-sky galaxy catalogue at this distance range, can be used for the detailed study of their origin (e.g., Pierre Auger Collaboration 2010; He *et al.* 2016).

2.5.4.5 Applications in transient astronomy

In the following paragraphs we outline potential applications of the *HECATE* in various other fields of transient and multi-messenger astrophysics.

Tidal disruption events (TDEs) are typically witnessed as an outburst in X-ray or optical wavelengths resulting from accretion of the material shredded from a star under the effect of the tidal field of a supermassive black-hole (SMBH). Such events are expected to be routinely detected in the *eROSITA* all sky X-ray survey, and the *LSST* optical survey. *HECATE* can provide the basis for the quick identification of the host of such an event and its basic properties. In particular, information on the distance, the presence of an AGN, and the velocity dispersion (used to initially estimate the SMBH mass; all available in the *HECATE*) are valuable for a quick interpretation of transient events (e.g., French *et al.* 2020).

A value-added catalogue providing robust distances and stellar population parameters is also useful for the characterisation and study of the populations of transient

events observed in on-going or future multi-wavelength surveys. For example, *LSST* is expected to provide a host of supernovæ every night. The association of these events with a catalogue like *HECATE* will facilitate systematic studies of their populations in the context of their host galaxies (e.g., $M_\star$, SFR, and most importantly metallicity; e.g., Greggio & Cappellaro 2019). These pilot studies can be used to effectively plan more focused follow-up observations. The same holds for the identification of hosts of fast radio bursts (e.g., Marcote *et al.* 2020).

2.6 Conclusions and future work

We present a new catalogue of galaxies which includes all known galaxies within a distance limit of $D \lesssim 200$ Mpc. We

1. use all available distance measurements for the sample to get robust redshift-independent distances, which are preferred over recessional velocity based estimates for galaxies in the local Universe, for as many galaxies as possible (10%),
2. compute redshift-dependent distances for the rest of the galaxies (90%) which are consistent with the redshift-independent distances (Kernel Regression method), while quantifying their uncertainties due to the unknown peculiar velocity component,
3. incorporate integrated multi-band photometry with special treatment for nearby and/or extended galaxies,
4. derive SFRs, $M_\star$, metallicities, and nuclear activity classifications utilising the best available information for each galaxy,
5. offer five different IR-based SFR indicators, as well as, a homogenised SFR indicator, while providing all the necessary information for user-defined calibrations.

Despite its limitations in terms of the completeness of the catalogue (§2.5.2), and data coverage (§2.5.3), the *HECATE* is the most complete sample of known galaxies in the local Universe. Owing to its wealth of information, the *HECATE* can be a useful tool for a wide range of applications. By providing positions and size information the catalogue can be used as the basis of future associations of galaxies with additional multi-wavelength surveys. We discuss a wide range of applications, including the prioritisation of host galaxies for follow-up searches for EM counterparts of GW

sources, as well as, the initial characterisation of transient sources which will be critical in the era of Big Data of astronomy.

Future versions of the catalogue will expand the distance range beyond the current limit of 200 Mpc, and provide a wider coverage in terms of the stellar population parameters. SFR and $M_\star$ estimates will be improved by: (i) including additional multi-wavelength data, (ii) adoption of forced-photometry techniques allowing the full exploitation of existing all-sky surveys also for extended objects, and (iii) performing SED analysis. Finally, incorporation of different sources of spectroscopic data will not only extend the coverage of metallicity and nuclear activity classifications, but more importantly will serve as a cross-validation dataset for AGN classifications. This is crucial for many areas of applications (e.g., screening for AGN in X-ray studies of galaxies, identification of candidate sources of high-energy γ-ray or cosmic rays).

This page intentionally left blank

Appendix

Appendix 2.A Computation of the Virgo-infall corrected radial velocities

Starting from the heliocentric velocity, v_{hc}, of a galaxy at galactic coordinates (l, b), we adopt the correction of Karachentsev & Makarov (1996) for solar motion in the Local Standard of Rest (LSR) and the Milky Way's motion with respect to the Local Group (LG) centroid:

$$v_{\rm lg} = v_{\rm hc} + V_a \left[\cos b \cos b_a \cos\left(l - l_a\right) + \sin b \sin b_a\right], \tag{2.11}$$

where $V_a = (316 \pm 5)$ km s^{-1} is the velocity of the Sun towards the LG centroid at galactic coordinates $l_a = (93 \pm 2)^\circ$ and $b_a = (-4 \pm 2)^\circ$. Then, we correct for the Local Group's infall to the Virgo cluster following Terry *et al.* (2002):

$$v_{\rm vir} = v_{\rm lg} + V_{\rm lg-infall} \cos\Theta, \tag{2.12}$$

where $V_{\rm lg-infall} = (208 \pm 9)$ km s^{-1} is the infall velocity of LG to Virgo cluster, and Θ is the great-circle distance between the galaxy's supergalactic coordinates and LG's apex $(102.^\circ 88, -2.^\circ 34)$.

Appendix 2.B Cross-matching procedures and obtained data

The following paragraphs provide additional details regarding some of the cross-matching procedures described in §2.4.1.

2.B.1 *HyperLEDA* vs. *NED*

The cross-correlation of the *HyperLEDA* and the *NED* is an essential step to (i) obtain missing radial velocities, (ii) use the associations to match *HyperLEDA* objects to z-independent distance measurements in *NED-D*, and (iii) provide quick links to *NED* entries for the galaxies. This step of the pipeline is executed before applying the recession velocity cut, since *NED* complements our sample with radial velocities. Therefore, 884,766 objects are searched, i.e. galaxies with heliocentric velocity $<14,500\,\mathrm{km\,s^{-1}}$ (ensuring that no object with $v_{\rm vir}<14,000\,\mathrm{km\,s^{-1}}$ is excluded), and objects without radial velocity information in *HyperLEDA*. We use the Python 'astroquery' package to associate the *HyperLEDA* galaxies to *NED* objects on the basis of their designation: for each object in *HyperLEDA*, we perform two searches: based on their PGC ID (e.g., PGC000002) and principal designation (e.g., UGC12889). We perform a series of checks to identify cases where:

1. the two searches (principal designation and PGC number) return different *NED* objects (1,024).
2. different *HyperLEDA* objects are associated to the same *NED* object (510),
3. positions or radial velocities disagree (1,232),
4. the *HyperLEDA* object has a large astrometric error and size, and has been associated to a *NED* object by chance (usually Zone of Avoidance objects; 33,101),
5. there are typographic errors in galaxy pairs (e.g., A in the place of B in *NED*; 202)

The above situations are resolved automatically (e.g., positional disagreement larger than 1 arcmin), or after manual inspection. In total, 137,586 galaxies (67%) are associated to *NED* objects.

2.B.2 Supplementary size information

HyperLEDA provides the size of the galaxies based on the D_{25} isophote in the B-band. However, for 39,251 objects (19%) this information is not available. Using the associations of *HyperLEDA* to *NED* objects, we find that for the majority of these objects, the diameters can be obtained from *2MASS* and *SDSS*. In addition, using the CDS XMatch service (http://cdsxmatch.u-strasbg.fr/), we find eight other catalogues that can provide diameters for the majority of the rest of these objects. The catalogues used to draw this information are listed in Table 2.B.1.

The supplementary size information is incorporated by rescaling the semi-major axis from the external catalogue, $a_{\rm ext}$, using as reference the *HyperLEDA* semi-major axis, $a_{\rm hyp}$. To do so, we:

1. associate all *HyperLEDA* objects to the external catalogue,

2. use the associated galaxies for which both $a_{\rm hyp}$ and $a_{\rm ext}$ are defined to compute the scaling factor $c=\langle a_{\rm hyp}/a_{\rm ext}\rangle$, and

3. fill in the $a_{\rm hyp}$ for the galaxies in *HECATE* without semi-major axis from *HyperLEDA*: $a_{\rm hyp}=c\times a_{\rm ext}$.

The priority of the external catalogues was selected on the basis of the quality of the scaling relation between the proposed diameter and the D_{25} in the *HyperLEDA* for the common objects (e.g., number of common of objects, scatter) and the proximity of the band they use to the B-band. More details can be found in Table 2.B.1.

When available, semi-minor axes and position angles are also taken from the external catalogues (the axis ratio in the *HECATE* is the same as the one reported by the external catalogue). In total, we complete the size information for 34,413 galaxies, leaving 4,837 (2.4%) galaxies without such information in the *HECATE*. Because of the different wavebands and methods used by the external catalogues, the application of a scaling factor is over-simplistic and may have introduced biases. Users of the catalogue are suggested to use the corresponding flag, 'dsource', to either filter out these galaxies, or study any biases.

Table 2.B.1: The various sources of semi-major axis information incorporated in the *HECATE*. Where available, axis ratios and position angles are also obtained. The semi-major axes are rescaled to match the D_{25} isophotal one in *HyperLEDA*. The columns are: (1) the name of the source; (2) The flag in the column 'dsource' in the provided catalogue; (3) the number of objects in the *HECATE* for which the sizes where obtained from the source; (4) the scaling factor C in dex, used to homogenise the sizes r_1 from the given source to R_1 (adopted semi-major axis) as in $\log R_1 = \log r_1 + C$; (5) the scatter (in dex) between the size in *HyperLEDA* and the source for the common objects; (6) notes; and reference of the source.

Source (1)	Flag (2)	Number (3)	Scale (4)	Scatter (5)	Notes & Reference (6)
HyperLEDA	H	165,482			This is the reference sample. Adopted as they are. Cf. (Makarov *et al.* 2014)
SDSS	S	12,214	0.208	0.188	Petrosian radius in the g-band from Data Release 15. The g-band was selected because of the small scatter in the scaling factor, as expected due to its proximity to the B-band. Cf. Aguado *et al.* (2019)
2MASS	2	12,918	0.236	0.118	Super-coadd 3-sigma isophotal semi-major axis radius ('sup_r_3sig'). The J-band 21 mag/arcsec^2 isophotal semi-major axis presents a slightly small scatter of 0.115 but it was not available for all objects. Cf. Jarrett *et al.* (2000)
6dFGS	6	6,327	1.940	0.065	The scaling factor converts from pixels to arcmin. Cf. Jones *et al.* (2004)
WINGS	W	740	-0.056	0.104	Cf. Moretti *et al.* (2014)
SkyMapper	Y	1,814	0.355	0.203	Data Release 1.1. Cf. Wolf *et al.* (2018)
AMIGA-CIG	A	65	-0.255	0.137	Cf. Verdes-Montenegro *et al.* (2005)
UNGC	K	60	-0.068	0.170	Cf. Karachentsev *et al.* (2013)
VIII/77	V	28	-0.069	0.121	Catalogue in the Vizier service. Cf. Springob *et al.* (2005)
KKH2001	1	26			No correction applied (B-band isophotes). Cf. Karachentsev *et al.* (2001)
KKH2007	7	9			No correction applied (B-band isophotes). Cf. Karachentsev *et al.* (2007)
NED	N	212			No correction applied. Miscellaneous diameters based on B-band, mainly from *ESO-LV*. See http://ned.ipac.caltech.edu

2.B.3 *IRAS-RBGS*

We cross-match objects in the *HECATE* and *IRAS-RBGS* on the basis of their D_{25} elliptical regions. 589 galaxies out of the 629 objects in *IRAS-RBGS* are associated to *HECATE* galaxies. The remaining 40 objects are not cross-linked for the following reasons. 19 associations are rejected because they are galaxy pairs that are resolved in the *HECATE* but unresolved in *IRAS-RBGS*: AM1633-682, ESO 60-IG016, ESO 255-IG007, ESO 343-IG013, IC 0563/4, IC 2810, IC 4518A/B, NGC 3395/6, NGC 3994/5, NGC 4038/9, NGC 4568/7, NGC 4922, NGC 5394/5, NGC 6052, NGC 6670A/B, NGC 7592, NGC 7752/3, NGC 5257/8, and UGC 12914/5. In addition, 21 *IRAS-RBGS* objects are not found in the *HECATE* because: (i) their radial velocity exceeds our recession velocity limit $v_{\rm vir}$=14, 000 km s^{-1} (18 galaxies), (ii) their object type in *HyperLEDA* is unknown (ESO 221-IG010 and ESO 350-IG038), or (iii) is identified as a star (IRAS F05170+0535).

2.B.4 *RIFSCz*

Before the cross-matching of *RIFSCz* and *HECATE*, we corrected an object designation in *RIFSCz* which was appearing twice in the catalogue (column 'FSCNAME'): two instances of F14012+5434, one of which was corrected to F01339+1532, after manual inspection using the provided coordinates. We associate the *HECATE* objects (without associations to *IRAS-RBGS*) to *RIFSCz* objects, if the D_{25} elliptical region of the former and the 6 arcsec circle (i.e., the resolution of *IRAS*) of the latter overlap. We find thousands of multiple matches. In order to resolve the multiple matches, we apply a four-step procedure:

1. Since *RIFSCz* provides better positional accuracy than *IRAS* (through associations to other surveys such as *2MASS*), we use a matching radius of 3 arcsec to cross-link the *HECATE* and *RIFSCz*. 18,147 matches are accepted, after resolving manually six multiple matches on the basis of radial velocities and offsets of the matched sources.
2. For the objects in *HECATE* and *RIFSCz* that remain unmatched after step (i), we use a 6 arcsec match radius for both catalogues. We find 550 matches, after resolving manually six multiple matches with the same criteria as in (i).
3. The unmatched objects (after (i) and (ii)), are cross-linked using the D_{25} region in the *HECATE* and the 6 arcsec circle around the position in the *RIFSCz*. Multiple matches are resolved with the requirement that radial velocities match

(<100 km s^{-1} difference). 407 matches are found, leaving only 168 unmatched objects in the *HECATE*, and 175 in *RIFSCz*.

4. The matches of the steps (i)-(iii) are joined and inspected for ambiguous matches, i.e., galaxy pairs may be resolved in the *HECATE* but not in the *RIFSCz*. We reject 22 such associations.

The above steps provide 19,082 unique associations between *HECATE* and *RIFSCz* objects.

2.B.5 *2MASS*

We sequentially cross-match the *HECATE* with the three catalogues providing *2MASS* data: *2MASS-LGA*, *2MASS-XSC* and *2MASS-PSC*. This order ensures that the associated photometric data reflect the full extent of the galaxies.

Out of the 665 objects in *2MASS-LGA*, we exclude 35 because they are not galaxies (see https://irsa.ipac.caltech.edu/data/LGA/overview.html). Out of the remaining 620 galaxies, 609 are cross-matched to *HECATE* objects. The unassociated objects were either exceeding the radial velocity criterion (7 objects), or *HyperLEDA* did not classify them as galaxies (3 objects), or belonged to the galaxy pair Arp 244 that is resolved in the *HECATE* (1 object).

Then, we cross-match the *HECATE* and the *2MASS-XSC* using a 3 arcsec match radius. From this procedure, we exclude the *HECATE* objects that are already associated to the *2MASS-LGA* galaxies. We also exclude the following extended galaxies that are resolved in the *2MASS*, and would produce thousands of chance coincidence matches: Draco Dwarf, Leo B, Sextans Dwarf Spheroidal, the Magellanic Clouds and Carina Dwarf Spheroidal. In total, we find 117,713 matches.

Finally, considering objects not associated to either the *2MASS-LGA* or the *2MASS-XSC*, we cross-match *HECATE* and *2MASS-PSC* and find 25,224 matches.

Appendix 2.C Empirical formulÆ for the distances of the *HECATE* galaxies

The intrinsic distance modulus $\mu_{\rm int}$ and its uncertainty $\epsilon_{\rm int}$ of a galaxy with Virgo-infall corrected radial velocity $v_{\rm vir}$, inferred by the Kernel Regression explained in

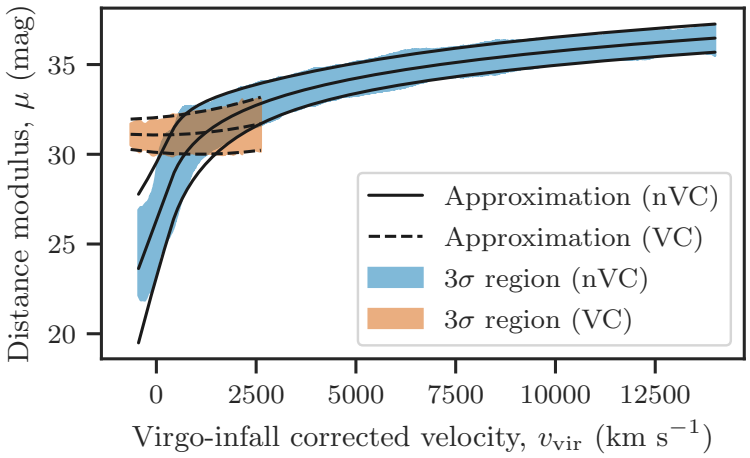


Figure 2.C.1: The local 99.7% intervals of the distance modulus of the two models, orange for the Virgo cluster and blue for the rest. The black lines correspond to the mean and same interval (dashed for VC model) computed using the approximation formulæ (Appendix 2.C).

§2.3.2.2, can be approximated by the following formulæ for nVC galaxies:

$$\mu_{\rm int} \approx \begin{cases} 26.34 + 0.006057u, & u \leq 358.5 \\ 15.74 + 5\log_{10} u, & u > 358.5 \end{cases}, \tag{2.13}$$

$$\epsilon_{\rm int} \approx 0.2611 + 0.8016 \exp\left(-\frac{u}{1341}\right), \tag{2.14}$$

where $u \equiv v_{\rm vir}/\,{\rm km\,s^{-1}}$. The above relations are valid for the range $u \in [-481.7, 14,033.0]$. Similarly, for VC galaxies:

$$\mu_{\rm int,VC} \approx 31.08 + 9.177 \times 10^{-8} u^2, \tag{2.15}$$

$$\epsilon_{\rm int,VC} \approx 0.3235 + 6.464 \times 10^{-5} u, \tag{2.16}$$

valid in the range $u \in [-792.5, 2764]$. These approximating formulæ for the distance modulus $\mu_{\rm int}$, and the $3\epsilon_{\rm int}$ region are plotted in Figure 2.C.1, on top of the corresponding quantities computed using the regression models.

Appendix 2.D Description of columns in the provided catalogue

The columns of the *HECATE* are described in Table 2.D.1.

Table 2.D.1: Description of the columns in the machine-readable catalogue. In cases of adopted values from external catalogues, the middle column reports the source: H=*HyperLEDA*, N=*NED*, I=*IRAS*, F=*WISE* forced photometry, M=*2MASS*, S=*SDSS*, G=*GSWLC-2*. Unflagged columns were computed by us.

Column		Description
pgc, objname	H	*Principal Galaxy Catalogue* number, and name in the *HyperLEDA*.
id_ned, id_nedd	N	Name in *NED* and *NED-D* respectively.
id_iras	I	Name in *IRAS-RBGS*, or in *RIFSCz* if in the form Fxxxxx+xxxx.
id_2mass	M	ID in *2MASS-LGA*, *2MASS-XSC*, or *2MASS-PSC* (see flag_2mass).
sdss_photid, sdss_specid	S	*SDSS* photometric and spectroscopic IDs (consistent with DR >8).
ra, dec	H	Decimal J2000.0 equatorial coordinates (deg).
f_astrom	H	Astrometric precision flag. -1 for ∼0.1 arcsec; 0 for ∼1 arcsec; 1 for ∼10 arcsec; and so on.
r1, r2, pa	H	D_{25} semi-major and semi-minor axes (arcmin), and North-to-Northeast position angle (deg).
rsource, rflag		Source (see Table 2.B.1) and flag of the size information: 0=missing, 1=all size information defined, 2=either r2 or pa were missing and they were set equal to r1 and 0.0 respectively (circular isophote).
t, e_t	H	Numerical Hubble-type and its uncertainty. See de Vaucouleurs *et al.* (1976).
incl	H	Inclination (deg).
v, e_v	HN	Heliocentric radial velocity, and its uncertainty (km s^{-1}).
v_vir, e_v_vir		Virgo-infall corrected radial velocity and its uncertainty (km s^{-1}).
ndist		Number of distance measurements in NED-D used for the computation of d.
edist		If True, the *NED-D* distance measurements used had uncertainties.
d, e_d		Distance, and it uncertainty (Mpc).
d_[lo/hi][68/95]		68% and 95% confidence intervals on the distance.

Continued on next page ...

Table 2.D.1 – *Continued from previous page*

Column		Description
dmethod		Method for the estimation of the distance: N=from *NED-D*, Z=regressor, Zv=VC-regressor, C(v)=distance from *NED-D* but uncertainty from (VC-)regressor.
ut, bt, vt, it	H	Total U, B, V, and I magnitudes (mag).
e_[u/b/v/i]t	H	Uncertainties on ut, bt, vt, it (mag).
ag, ai	H	Galactic and intrinsic absorption in *B*-band.
s12, s25, s60, s100	I	*IRAS* fluxes at 12, 25, 60, and 100μm respectively (Jy).
q12, q25, q60, q100	I	Quality flags for s12, s25, s60, s100: 0=not in *IRAS*, 1=upper limit, 2=moderate quality, 3=high quality in *FSC* or 4=flux from *RBGS*.
wf1, wf2, wf3, wf4	F	3.3, 4.6, 12 and 22μm fluxes in the *WISE* forced photometry catalogue (mag).
e_wf1/2/3/4	F	Uncertainties on wf1, wf2, wf3, wf4 (mag).
wfpoint, wftreat	F	'True' if point source, and 'True' if treated as such, respectively, in the *WISE* forced photometry catalogue.
j, h, k	M	J, H, and K_s-band apparent magnitudes in *2MASS* (mag).
e_j, e_h, e_k	M	Uncertainties on j, h, k (mag).
flag_2mass		Source of the *2MASS* ID and JHK magnitudes: 0=none, 1=*LGA*, 2=*XSC*, 3=*PSC*.
u, g, r, i, z	S	u, g, r, i, and z-band apparent magnitudes in *SDSS* (mag).
e_u/g/r/i/z	S	Uncertainties on u, g, r, i, z (mag).
logL_TIR		Decimal logarithm of the TIR luminosity ($L_\odot$).
logL_FIR		Decimal logarithm of the FIR luminosity ($L_\odot$).
logL_60u		Decimal logarithm of the 60μm-band luminosity ($L_\odot$).
logL_12u		Decimal logarithm of the 12μm-band luminosity ($L_\odot$).
logL_22u		Decimal logarithm of the 22μm-band luminosity ($L_\odot$).
logL_K		Decimal logarithm of the K_s-band luminosity ($L_\odot$).
ML_ratio		Mass-to-light ratio (§2.4.2.1).
logSFR_TIR		Decimal logarithm of the TIR-based SFR estimate ($M_\odot\,yr^{-1}$).
logSFR_FIR		Decimal logarithm of the FIR-based SFR estimate ($M_\odot\,yr^{-1}$).
logSFR_60u		Decimal logarithm of the 60μm-based SFR estimate ($M_\odot\,yr^{-1}$).
logSFR_12u		Decimal logarithm of the W3-based SFR estimate ($M_\odot\,yr^{-1}$).
logSFR_22u		Decimal logarithm of the W4-based SFR estimate ($M_\odot\,yr^{-1}$).

Continued on next page ...

Table 2.D.1 – *Continued from previous page*

Column		Description
logSFR_HEC		Homogenised log SFR ($M_\odot\ yr^{-1}$). Rescaling of SFR indicators is performed only here (§2.4.2.2).
SFR_HEC_flag		Flag indicating photometry source and SFR indicator used for logSFR_HEC (Table 2.2).
logM_HEC		Decimal logarithm of the $M_\star$ ($M_\odot$).
logSFR_GSW	G	Decimal logarithm of the SFR in *GSWLC-2* ($M_\odot\ yr^{-1}$).
logM_GSW	G	Decimal logarithm of the $M_\star$ in *GSWLC-2* ($M_\odot$).
min_snr		Minimum signal-to-noise ratio among all emission lines used for the activity classification (class).
metal, flag_metal		Metallcity $[12 + \log(O/H)]$ and its quality flag (§2.4.2.3).
class_sp		Nuclear activity classification (§2.4.2.4): 0=star forming, 1=Seyfert, 2=LINER, 3=composite, -1=unknown.
agn_s17	E	AGN classification in She *et al.* (2017): Y=AGN, N=non-AGN, ?=unknown.
agn_hec		Combination of SDSS and She *et al.* (2017) classifications (§2.4.2.4): Y=AGN, N=non-AGN, ?=unknown.

3

A census of ultraluminous X-ray sources in the local Universe

3.1 Background

Ultraluminous X-ray sources (ULXs) are galactic point-like X-ray sources, not associated with an active galactic nucleus, with X-ray luminosities above the Eddington limit of an accreting stellar-mass black hole ($\gtrsim 10^{39}$ erg s^{-1}; for a recent review see Kaaret *et al.* 2017). Soon after their discovery by the *Einstein* observatory (Long & van Speybroeck 1983; Fabbiano 1989), three scenarios were proposed to explain their high luminosities. Initially, it was proposed that ULXs are accreting black holes (BHs) with masses in the range between stellar-mass and supermassive BHs ($\sim 10^2 - 10^6$ M$_\odot$), i.e., intermediate-mass BHs (IMBHs; Colbert & Mushotzky 1999; Makishima *et al.* 2000; van der Marel 2004). This scenario was dismissed on theoretical grounds due to difficulties in the formation of X-ray binaries with IMBHs (e.g., Kuranov *et al.* 2007), although a few cases are still viable (e.g., ESO 243-49 HLX-1: Farrell *et al.* 2009; M82 X-1: Ptak & Griffiths 1999). The second scenario involves stellar-mass BHs (with masses in

the range of Galactic BHs, $\lesssim$15 $M_\odot$; Remillard & McClintock 2006), which may have super-Eddington luminosities when accreting at super-critical rates (e.g., Begelman 2002). In the third scenario the ULX luminosities are the result of geometrical beaming of the emitted radiation (King *et al.* 2001) due to the formation of a funnel in the central part of the supercritical accretion disk (e.g., Abramowicz *et al.* 1988; Sądowski *et al.* 2014b).

The combination of these two scenarios can explain the observed ULX population with L_X<10^{41} erg s^{-1} as the high-luminosity end of the luminosity function of X-ray binaries (XRBs). Recently, the discovery of pulsating ULXs (Bachetti *et al.* 2014; Fürst *et al.* 2016; Israel *et al.* 2017a,b; Carpano *et al.* 2018) showed that the accretor can even be a neutron star (NS), making the super-Eddington accretion scenario necessary for their explanation (e.g., Fragos *et al.* 2015; King & Lasota 2016; King *et al.* 2017; Middleton & King 2017; Misra *et al.* 2020).

The above three scenarios highlight the importance of ULXs in understanding massive binary evolution and accretion physics at extreme accretion rates. The latter is crucial for shedding light at the formation of compact object mergers that are detected as short gamma-ray bursts and gravitational wave sources (Berger 2014; Finke & Razzaque 2017; Marchant *et al.* 2017; Mondal *et al.* 2020). In addition, the extreme emission of ULXs may have played a role in the heating of the Universe during the epoch of reionization (e.g., Venkatesan *et al.* 2001; Madau *et al.* 2004; however see Das *et al.* 2017; Madau & Fragos 2017).

A deeper understanding of ULXs can be obtained by detailed spectral and timing studies of individual sources (e.g., Gladstone *et al.* 2009; Middleton *et al.* 2015; Walton *et al.* 2019; Koliopanos *et al.* 2019). While these studies provide valuable insights into the physics and nature of the accretion, they offer limited information on the formation and evolution pathways of ULXs. The latter can be better constrained by identifying their optical counterparts and/or studying their populations in the context of their host galaxies. Since ULXs are rare and usually found in distant galaxies, the identification of optical counterparts and measurement of the compact object masses are observationally challenging (Angelini *et al.* 2001; Colbert & Ptak 2002; Swartz *et al.* 2004; Feng & Kaaret 2008). Consequently, ULX demographics and scaling relations between the ULX content and stellar population parameters of their host galaxies, such as SFR and $M_\star$, are important tools for understanding the nature and evolution

of ULXs via the comparison with binary population synthesis models (e.g., Rappaport *et al.* 2005; Wiktorowicz *et al.* 2017).

Early surveys of nearby galaxies revealed an overabundance of ULXs in late-type galaxies (LTGs) (e.g., Roberts & Warwick 2000), while direct association of ULXs with star-forming regions of their hosts connected ULXs with young stellar populations, indicating that the majority of ULXs are a subset of high-mass X-ray binaries (HMXBs; e.g., Fabbiano *et al.* 2001; Roberts *et al.* 2002; Gao *et al.* 2003; Zezas *et al.* 2007; Wolter & Trinchieri 2004; Kaaret *et al.* 2004; Anastasopoulou *et al.* 2016; Wolter *et al.* 2018). Nevertheless, a small but significant fraction of ULXs are found in early-type galaxies, and therefore are connected to old stellar populations, i.e. ultraluminous low-mass X-ray binaries (LMXBs; Angelini *et al.* 2001; Colbert & Ptak 2002; Swartz *et al.* 2004; Kim & Fabbiano 2004; Fabbiano *et al.* 2006; Feng & Kaaret 2008. These demographic studies agree on two findings:

a) Dwarf galaxies have been found to host more ULXs than expected given their SFR (Swartz *et al.* 2008; Walton *et al.* 2011; Plotkin *et al.* 2014; Tzanavaris *et al.* 2016).

b) An observed excess of ULXs (and XRBs in general) in low-metallicity galaxies (e.g., Mapelli *et al.* 2010; Prestwich *et al.* 2013; Brorby *et al.* 2014; Douna *et al.* 2015; Basu-Zych *et al.* 2016)

The excess in low-metallicity galaxies has highlighted the effect of metallicity on the accretor's mass and the evolutionary paths of ULXs (Heger *et al.* 2003; Soria *et al.* 2005; Belczynski *et al.* 2010; Linden *et al.* 2010; Mapelli *et al.* 2011; Marchant *et al.* 2017). The same effect has been invoked to interpret the X-ray emission properties of high-redshift galaxies (Lyman Break Galaxies and Lyman Break Analogs; Basu-Zych *et al.* 2013a,b, 2016; Brorby *et al.* 2016; Lehmer *et al.* 2016), as demonstrated by binary population synthesis models (Linden *et al.* 2010; Fragos *et al.* 2013a,b; Wiktorowicz *et al.* 2017).

In the era of *ROSAT* and the early days of *Chandra*, ULX demographics were limited to a few tens of sources and galaxies (e.g., Colbert & Ptak 2002; Swartz *et al.* 2004; Liu *et al.* 2006). Therefore, these studies were unable to resolve the dependence of ULX populations on the stellar populations of their host galaxies. The first quantitative study of the rate of ULXs in the local Universe, based on a complete sample of galaxies up to 14.5 Mpc, showed that the observed population of ULXs is 'consistent with the extrapolation of the luminosity function of ordinary X-ray binaries' (LMXBs and HMXBs

in early- and late-type galaxies respectively; Swartz *et al.* 2011). However, the volume limit resulted into an oversampling of irregular galaxies and under-representation of elliptical galaxies. The largest to date demographic study of ULXs (343 galaxies) was presented in Wang *et al.* (2016), using *Chandra* observations until 2007. This work constrained the X-ray luminosity function (XLF) parameters of ULXs in galaxies of different morphological types, and showed that elliptical galaxies host more ULXs than in samples of previous studies. However, Wang *et al.* (2016) focused on XLFs of ULXs and did not study their scaling with the SFR, $M_\star$ and metallicity of their hosts. The most recent catalogue of ULX candidates was presented in Earnshaw *et al.* (2019). It includes 384 ULXs drawn from the 3XMM-DR4 catalogue. This study showed that the hardness ratio distribution of ULXs is similar to that of the lower-luminosity XRBs, but not AGN, and mostly independent of the environment (elliptical vs. spiral galaxies). However, this study focused on the X-ray spectral and timing properties of the sources rather than their connection to their hosts.

The *Chandra Source Catalog 2.0* (*CSC 2.0*) gives a unique opportunity to study the demographics of ULXs in the context of the stellar populations of their host galaxies (SFR, $M_\star$, metallicity) by utilising the largest available sample of X-ray sources, and a new catalogue of galaxies in the local Universe.

This paper is organised as follows: in Sections 3.2 and 3.3 we describe the sample of host galaxies and X-ray sources, respectively. In Section 3.4 we report the results on ULX demographics and their connection with stellar population parameters, while in Section 3.5 we discuss the implications of this study in comparison to previous studies and ULX population models. Finally, in Section 3.6 we summarise the main findings. Unless stated otherwise, the reported uncertainties correspond to 68% confidence intervals.

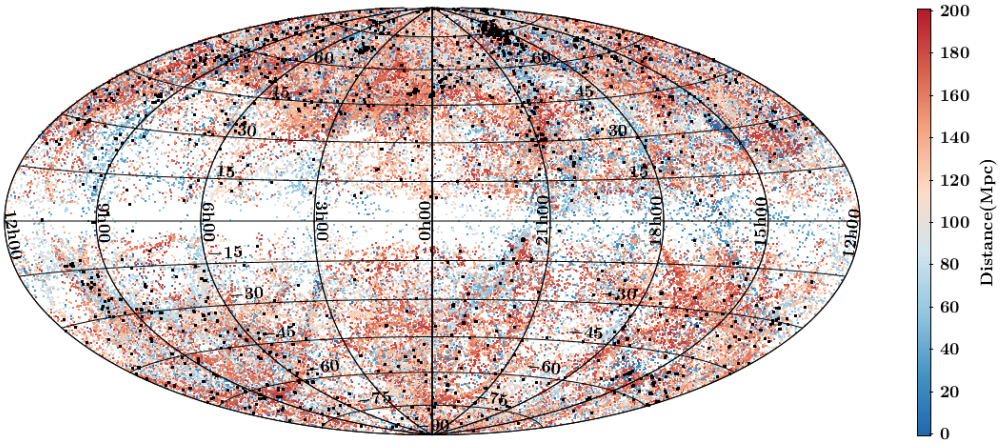


Figure 3.1.1: Sky map of the *HECATE* galaxies in Galactic coordinates with colour denoting their distance. Galaxies included in the *CSC 2.0* are shown as black points. Note the sparsity of sources in the plane of the Milky Way (Zone of Avoidance) and the increased density in the North and West parts due to the inclusion of *SDSS* galaxies in the *HyperLEDA*.

3.2 The galaxy sample

We use the *Heraklion Extragalactic Catalogue* (*HECATE*), a compilation of all galaxies within 200 Mpc, from the *HyperLEDA* (Makarov *et al.* 2014), arguably the most complete compilation of galaxies with homogenised parameters. The *HECATE* adopts positions, sizes, morphological classifications, and redshifts from the *HyperLEDA*. These are complemented with size and redshift information from other catalogues when not available in the *HyperLEDA*. It also provides robust estimates of distances, along with SFRs, stellar masses, metallicities and nuclear activity classifications. In the following paragraphs we provide a brief summary of the relevant properties of the catalogue. A detailed description of the catalogue and the data it contains is presented in §2.

The *HECATE* is based on all *HyperLEDA* galaxies (object type 'G') with Virgo-infall corrected radial velocities less than 14000 km s^{-1} (corresponding to distances $\lesssim$200 Mpc and redshifts $\lesssim$0.047). When redshift and size information (semi-major/minor axes and position angles) are not directly available in the *HyperLEDA*, they are obtained from other databases or catalogues (e.g., *NED*, *SDSS*, *2MASS*)*. Figure 3.1.1 shows the

* None of these galaxies (with supplemented redshift/size information) is included in our analysis because they lack other required information (e.g., morphological classifications, IR photometry which is used for deriving SFR and stellar mass measurements).

Table 3.2.1: Parameters of the 2218 host galaxies. Only a small portion of this table is shown here, indicative of the various cases (e.g., flags, missing parameters). The full table is available in the online journal.

PGC	ID	α	δ	R_1	R_2	ϕ	T	D	log SFR	log $M_\star$	Z	AGN	f_{25}	U	$N_{\rm obs}$	$N_{\rm f/b}$	$N_{\rm ulx}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
2557	NGC0224	10.684684	41.268978	88.91	34.83	35	3.0±0.4	0.8	-0.33	10.61		Y	0.46		0	0.01	$0.00^{+1.13}_{-0.00}$
16570	NGC1741B	75.398106	-4.263220	0.46	0.23	42	6.8±3.3	55.7		9.34			1.00	*	2	0.09	$1.91^{+1.85}_{-1.13}$
23324	UGC04305	124.768125	70.721674	3.96	2.79	15	9.9±0.5	3.4	-2.07	8.77			0.27	*	1	0.00	$1.00^{+1.49}_{-0.73}$
35249	NGC3683	171.882672	56.877021	0.87	0.35	124	4.8±0.7	33.3	0.94	10.81	8.76	Y	1.00	*	5	0.12	$4.88^{+2.62}_{-1.93}$
38742	NGC4150	182.640252	30.401578	0.99	0.66	148	-2.1±0.7	13.6	-0.84	9.90		N	1.00	*	2	0.02	$1.98^{+1.84}_{-1.14}$

Column descriptions: (1) identification number in the *HyperLEDA* and the *HECATE*; (2) galaxy name; (3), (4) right ascension and declination (J2000.0) (°); (5)-(7) the semi-major and -minor axes (′), and the North-to-East position angle (°); (8) morphological code, T (see Table 3.3.2); (9) distance (Mpc); (10) decimal logarithm of SFR [$\mathrm{M_\odot\,yr^{-1}}$]; (11) decimal logarithm of $M_\star$ [$\mathrm{M_\odot}$]; (12) metallicity ($12 + \log(O/H)$); (13) the galaxy hosts an AGN; (14) fraction of D_{25} covered by the *CSC 2.0* stacks; (15) * if the galaxy is used in the analysis in this paper (see §3.3.2); (16) number of observed sources with $L_X{>}10^{39}$ erg s^{-1}, excluding nuclear sources if the host is classified as AGN (see §3.3.7); (17) number of expected foreground/background source contamination in the ULX regime; (18) number of ULXs by subtracting interlopers. Columns (1)-(13) are taken from the *HECATE*, while the rest are described in §3.3.

position of the galaxies in the *HECATE* in Galactic coordinates.

The *HECATE* provides redshift-independent distances (e.g., based on the Cepheids, RR Lyrae, Tully-Fisher, surface-brightness fluctuations, tip of the red-giant branch methods) for ∼10% of the galaxies obtained from the *NED-D* (Steer *et al.* 2017). When only one distance measurement is available, it is adopted as is. In the case of multiple distance measurements, a statistical estimate is made using a weighted Gaussian Mixture model, with weights that penalise uncertain or old measurements. Subsequently, these distances along with the radial velocities for the same galaxies are used to train a Kernel Regression model which is the used to predict the radial-velocity based distance (and its uncertainty based on the intrinsic scatter) for all the other objects. More details on the method for calculating the galaxy distances are given in §2.

The *HECATE* provides SFR estimates for galaxies with reliable mid- and far-infrared photometric measurements from the *IRAS* and the *WISE*. Depending on the availability and quality of photometry, three different SFR indicators were computed based on *IRAS* photometry: (i) total-infrared (TIR; 24, 60 and 100μ calibrations of Dale & Helou 2002 and Kennicutt & Evans 2012), (ii) far-infrared (FIR; 60 and 100μ calibrations of Helou & Walker 1988 and Kennicutt 1998), and (iii) 60μ (calibrations of Rowan-Robinson 1999). Additionally, *WISE* photometry, obtained from the forced photometry catalogue of Lang *et al.* (2016) for galaxies in the *SDSS* footprint, is used to provide 12μ and 22μ-based SFR estimates (calibrations of Cluver *et al.* 2017). An 'adopted' SFR for each galaxy is obtained by homogenising the SFR indicators (using the TIR-based one as reference) and selecting for each galaxy the first available SFR estimate in the following order of preference: TIR, FIR, 60μ, 12μ, 22μ. It should be noted that Hα SFR indicators are more appropriate than infrared (IR) indicators for studying the connection of ULXs with young stellar populations, since the latter probe star formation at scales (∼100 Myr), longer than the life-time of HMXBs (cf. Kouroumpatzakis *et al.* 2020). However, IR photometry is readily available for a significant fraction of our sample, and it is generally well correlated with Hα.

The integrated *2MASS* K-band photometry and *SDSS* $g-r$ colour, were used to estimate the stellar masses of the galaxies, using the mass-to-light ratio calibrations of Bell *et al.* (2003). For galaxies without *SDSS* photometry, the *HECATE* assumes the mean mass-to-light ratio of the galaxies with *SDSS* data.

In addition, the *HECATE* includes gas-phase metallicities based on *SDSS* spectroscopic data from the *MPA-JHU catalogue* (Kauffmann *et al.* 2003; Brinchmann *et al.*

2004; Tremonti *et al.* 2004), using the O III-N II calibration in Pettini & Pagel 2004. Based on the star-light subtracted *SDSS* spectra, the *HECATE* identifies AGN on the basis of their location in optical emission-line ratio diagnostic diagram, using the multi-dimensional classification scheme of Stampoulis *et al.* (2019).

The Third Reference Catalogue of Bright Galaxies (*RC3*; de Vaucouleurs 1991) has been the reference galaxy sample for several studies of ULXs (e.g. Swartz *et al.* 2011; Wang *et al.* 2016; Earnshaw *et al.* 2019). While it provides a wide range of information (positions, diameters, morphological types, photometry, and radial velocities), its small size (23022 galaxies) has been superseded by larger and more complete samples of galaxies. The *HyperLEDA*, and subsequently the *HECATE*, provide a ~10 times improvement in the sample size within our volume of interest (D<200 Mpc). Therefore, the *HECATE*, provides a much more complete census of the galaxy populations in the local Universe, supplemented by a wealth of additional information described in the previous paragraphs. This makes it more appropriate for the exploration of the multi-wavelength properties of galaxies based on serendipitous surveys.

3.3 The X-ray sample

To identify ULX candidates, we use the *CSC 2.0*[*], which is a publicly available catalogue of all the sources detected in *Chandra* observations performed up to the end of 2014. It contains 317167 X-ray sources, an improvement of more than a factor of 3 compared to the previous version (version 1.1; Evans *et al.* 2010).

3.3.1 Selection of sources

An X-ray source is associated with a *HECATE* galaxy if it is located within its D_{25} region. The positional uncertainties of the sources are not considered, since they are negligible with respect to the dimensions of the galaxies: 95% (98%) of the sources have uncertainties less than the 1% (10%) of the semi-major axes of their host galaxies. The few, galaxies without size information in the *HECATE* (~2% of the full sample) are excluded from the cross-matching.

Out of the 317167 sources in the *CSC 2.0*, we associate 23043 sources to 2218 galaxies within a distance of 200 Mpc. The host galaxies are shown by black points

[*] https://cxc.harvard.edu/csc2/

in Figure 3.1.1. The parameters of the host galaxies are listed in Table 3.2.1 (columns (1)-(13)), while the properties of the selected X-ray sources are given in Table 3.3.1.

We characterise sources as 'reliable' or 'unreliable' (column (9) in Table 3.3.1) based on their attributes in the *CSC 2.0*. A source is marked as 'unreliable' if any of the following conditions are met:

(i) the flux is zero (i.e., upper limit) or no confidence interval is provided,

(ii) the 'dither_warning_flag' is on: indicating that the highest peak of the power spectrum of the source occurs at the dither frequency (or its beat frequency) in all observations,

(iii) the 'streak_src_flag' is on: the source is found on an ACIS readout streak in all observations,

(iv) the 'sat_src_flag' is on: saturated in all observations.

We find 3783 (16.4%) 'unreliable' sources, out of which, 1040 (4.5%) are characterised as such because of a flag, 1952 (8.5%) have zero flux, and 791 (3.4%) have missing confidence intervals. The remaining 19260 sources (83.6%) are characterised as 'reliable'. In the following analysis we consider only the 'reliable' sources. We note that the majority of the more luminous sources in our sample ($L_X \gtrsim 10^{41.5}$ erg s^{-1}) are flagged as 'unreliable' (see §3.4.1), since they are more likely to be saturated.

3.3.2 *Chandra* field-of-view coverage

The *Chandra* observations from which the X-ray sources in the *CSC 2.0* are observed, typically target individual galaxies. The field of view is usually centred on the galaxy and covers fully its D_{25} region. However, there are cases of large nearby galaxies that are partially covered, as well as, observations that target off-centre regions.

In order to measure the coverage of each galaxy by *Chandra*, we compute the fraction, f_{25}, of the D_{25} region in the union of the stack-field-of-view* of all the stacks contributing in the *CSC 2.0* (column (11) in Table 3.2.1). We consider galaxies with $f_{25} > 0.7$ as sufficiently covered. After visually inspecting multi-wavelength images of the galaxies without full coverage, we find that the missing area generally leaves a negligible fraction of the total SFR and $M_\star$ unaccounted for.

* http://cxc.harvard.edu/csc2/data_products/stack/fov3.html

Table 3.3.1: Properties of the 23043 X-ray sources. Only a small portion of this table is shown here, indicative of the various cases (e.g., flags, missing parameters). The full table is available in the online journal.

PGC	ID	α	δ	$\log f$	$\log f_{\rm lo}$	$\log f_{\rm hi}$	p	u	n	c	$\log L_{\rm X}$	$\log L_{\rm X,lo}$	$\log L_{\rm X,hi}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
101	2CXO J000120.2+130641	0.33422	13.11141	-14.37	-14.54	-14.25				0.14	39.44	39.28	39.56
1305	2CXO J002012.6+591501	5.05281	59.25038	-13.36	-13.40	-13.32		*		0.97	36.47	36.42	36.50
2789	2CXO J004732.9-251748	11.88735	-25.29692	-12.15	-12.16	-12.15	*			0.17	39.02	39.01	39.03
12997	2CXO J032953.1-523054	52.47155	-52.51524	-13.87	-13.97	-13.80			*	0.02	40.60	40.51	40.68
42038	2CXO J123622.9+255844	189.09568	25.97891			-15.05		*		0.09			37.24

Column descriptions: (1) identification number of host galaxy in the *HyperLEDA* and the *HECATE*; (2) name of master source in the *CSC 2.0*; (3), (4) right ascension and declination (J2000.0) (°); (5)-(7) decimal logarithm of flux ('flux_aper90_b') and its 68% confidence interval [erg s^{-1} cm^{-2}]; (8) * if pileup source (lower limit on flux and luminosity); (9) * if 'unreliable' source (see §3.3.1); (10) * if nuclear source; (11) galactocentric scale parameter; (12-14) decimal logarithm of X-ray luminosity and its 68% confidence interval [erg s^{-1}]. The data in columns (2)-(8) are taken directly from the *CSC 2.0*, while those in columns (9)-(14) are described in §3.3.

Table 3.3.2: Morphological types and corresponding numerical codes (or indices) T, as described in the documentation of the Second Reference Catalogue of Bright Galaxies (de Vaucouleurs *et al.* 1976). Throughout this paper, we consider as early-type galaxies (ETGs) the elliptical ($T<-3.5$) and lenticular galaxies ($-3.5<T<-0.5$), and as late-type galaxies (LTGs) the rest. The morphological types of the galaxies are taken from the *HECATE*. Throughout the text, different binnings are described as ranges (e.g., Sdm-Im), and measurements with uncertainty less than 1.0 in T are considered reliable.

	Elliptical			Lenticular			Early spiral					Late spiral				Irregular (Irr)	
Morphological type	cE	E0	E+	$S0^-$	$S0^o$	$S0^+$	S0/a	Sa	Sab	Sb	Sbc	Sc	Scd	Sd	Sdm	Sm	Im
Numerical code, T	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
	ETGs						LTGs										

We find 34 galaxies (<2%) with coverage less than 70%. Some galaxies may have poor coverage because of observations performed in sub-array mode (e.g., those focusing on known ULXs.) Excluding these galaxies would bias our demographics; on the other hand, ULXs that are located in the unobserved area of the galaxies would also provide an incomplete picture of ULX populations. For this reason, we manually inspect for the presence of bright sources in *XMM-Newton* observations with wider field-of-view. Such observations are available for 16 objects, for which we find no other bright (L_X>10^{39} erg s^{-1}) sources in their D_{25} regions. Therefore, we include them in the following analysis since their ULX population is complete in our *Chandra*-based sample. The remaining 18 galaxies are excluded from the subsequent analysis (most of which are known to not host ULXs, e.g., SMC), but not from the provided catalogues.

3.3.3 Survey coverage and representativeness

3.3.3.1 Source confusion

At large distances, source confusion severely limits X-ray binary population studies. This effect is more prominent in studies of young stellar populations (such as ULXs; e.g., Anastasopoulou *et al.* 2016; Basu-Zych *et al.* 2016) due to the clumpy nature of star-forming regions (e.g., Elmegreen & Falgarone 1996; Sun *et al.* 2018). Specifically, at D>40 Mpc, the half-arcsecond beam of *Chandra* is comparable to the angular sizes of typical star-forming regions ($\lesssim$0.5 kpc; see discussion in Anastasopoulou *et al.* 2016). For this reason we restrict our analysis to the 644 galaxies in the host galaxy sample that are closer than 40 Mpc. This allows for direct comparisons with the works of Grimm *et al.* (2003) and Mineo *et al.* (2012) which adopt similar distance limits.

3.3.3.2 Observer bias

A limitation of this study is that our ULX sample is not based on a homogeneous, blind survey, but an accumulation of archival data gathered from targeted observations with different selection criteria. The unknown selection function may lead to observer biases, such as an over-representation of starburst galaxies: SFR is connected to the number of ULXs, as well as, other interesting phenomena (e.g. galaxy mergers) which may have been the focus of *Chandra* observations. To explore any biases or selection effects, Figure 3.3.1 shows the distributions of the (a) morphological types (see Table 3.3.2 for the morphological classification used in this paper), (b) SFRs, (c)

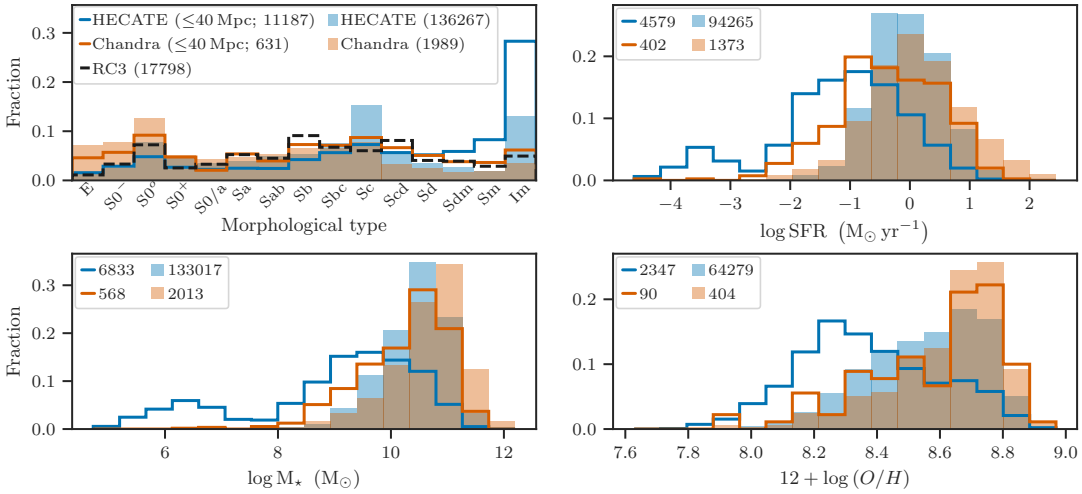


Figure 3.3.1: Distribution of morphological types (upper left), SFRs (upper right), $M_\star$ (lower left) and gas-phase metallicities (lower right) for the total volume of the *HECATE* (light blue bins) and the D<40 Mpc subset (blue steps), as well as, the *HECATE*/*CSC 2.0* galaxies (light orange bins) and their D<40 Mpc subset (orange steps). For reference, the distribution of morphological types in the *RC3* is shown with dashed black steps in the top left panel. The fractions are computed with respect to the sample size of each subset (shown in the legend).

stellar masses, and (d) metallicities for all galaxies with available relevant information in the *HECATE*. We compare the parent sample with the subset observed by *Chandra*, in the total volume and the D<40 Mpc limited sample.

In terms of morphology, *Chandra* has observed a slightly larger fraction of ETGs galaxies compared to late-type galaxies, a result of observations of nearby clusters which host larger populations of elliptical galaxies. In the D<40 Mpc sample, the *HECATE* includes a large population of irregular galaxies (mostly satellites of Local Group galaxies), though *Chandra* has observed only a small fraction of them. For comparison with previous works, in the top left panel of Figure 3.3.1 we show the distribution of the morphological types in *RC3*. The distribution is similar to the one of the host galaxy sample with D<40 Mpc. We also find that the distributions of $M_\star$, SFRs and metallicities of galaxies with *Chandra* observations in the total volume of the *HECATE*, are slightly shifted towards larger values than those in the parent sample.

These biases combined with the complex selection function of the *HECATE* and *Chandra* samples, do not allow us to calculate the volume density of ULXs. Nonetheless, the fact that the X-ray sample covers a wide range of SFRs, $M_\star$ and metallicities

characteristic of the local galaxies, allows us to draw representative scaling relations. Figure 3.3.2 shows the coverage in the SFR-$M_\star$ and sSFR-SFR planes for three different samples: the D<40 Mpc galaxies in the *HECATE*, the subset of those that are included in the *CSC 2.0*, and the subset of the latter hosting ULXs. We note that in this figure we exclude AGN-hosting galaxies to avoid biases in the stellar population parameters (see §3.3.7). We find that the host sample covers galaxies down to stellar mass of $10^{7.5}$ $M_\odot$ and SFR of $10^{-2.5}$ $M_\odot$ yr^{-1}, and is uniform in specific SFR (sSFR).

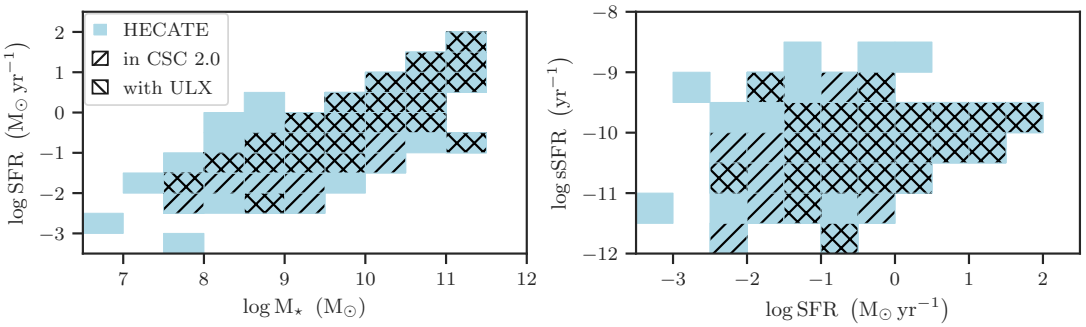


Figure 3.3.2: Coverage of the *HECATE* non-AGN galaxies with D<40 Mpc, in the SFR-$M_\star$ (left) and sSFR-SFR (right) planes. Right-angled hatching indicates galaxies covered by the *CSC 2.0*, while galaxies with left-angled hatching are ULX hosts. It is fairly representative for galaxies of $M_\star$>$10^{7.5}$ $M_\odot$ and SFR>$10^{-2.5}$ $M_\odot$ yr^{-1} and covers the full sSFR range in this region. ULX hosts (back diagonal hatching) cover the same parameter ranges, however, as expected, more sparsely in the low-SFR regime.

3.3.4 Galactocentric distances

The shape of the spatial distributions of ULXs in galaxies of different morphological types can provide valuable information regarding their association to the young or old stellar populations, and globular cluster systems. For this reason we calculate the galactocentric distance as the deprojected distance between the source and the centre of the galaxy, assuming the source resides in the disc of the galaxy. In order to normalise the measured galactocentric distance for the size of the galaxy, we derive the *galactrocentric scale parameter*, c (see Table 3.3.1), which we define as the ratio of the deprojected galactocentric distance of the source over the semi-major axis of the latter. A full description of the deprojection method and the calculation of c is presented in Appendix 3.A.

3.3.5 Source luminosities

In principle, spectral fitting is required for reliable estimates of the source fluxes. Due to the insufficient photon counts for most sources, we use the full-band (0.5 – 8.0 keV) aperture-corrected net energy flux inferred from the PSF 90% enclosed count fraction aperture as provided by the *CSC 2.0* (columns (5)-(7) in Table 3.3.1). In the case of sources with multiple observations, their fluxes are estimated from the 'longest observed segment based on a Bayesian Block analysis of all observations'*. We avoid the use of average fluxes (from coadds) since they systematically underestimate the flux of variable sources (e.g., Zezas *et al.* 2007). Indeed, we find that the above fluxes for our sources are, on average, ≈5% higher than their average* fluxes in the *CSC 2.0*.

We convert fluxes to luminosities (columns (12)-(14) in Table 3.3.1) adopting the distance of the host galaxy in the *HECATE* (column (6) in Table 3.2.1). The luminosities of 49 sources with significant pileup (column (8) in Table 3.3.1), are considered as lower limits. However, this does not affect the ULX demographics: their majority (41 sources) are excluded from the ULX demographics as nuclear sources in galaxies hosting an AGN (or without nuclear classification; see §3.3.7). Three of them are found in poorly-covered galaxies (excluded from our analysis; see §3.3.2) which are known to not host ULXs (SMC, LMC, Draco Dwarf). The remaining five piled-up sources present luminosities $>10^{39}$ erg s^{-1} and therefore are bona-fide ULX candidates, but their small number does not bias the luminosity distributions presented in this paper, while they are fully accounted for in the demographics.

3.3.6 Foreground/background contamination

The main source of contamination in large-area surveys are background (e.g. AGN) and foreground (e.g. stars) objects. Even through we cannot classify individual X-ray sources in our sample as ULXs, AGN, or other classes, using statistical techniques we can remove the effects of these contaminants from our analysis. As a first step we quantify the expected number of foreground/background (f/b) sources in the *CSC 2.0* footprint for each galaxy.

There are two commonly used methods to estimate the surface density of interlopers, based on: (i) blank fields around the target galaxies (e.g., Wang *et al.* 2016),

* see 'flux_aper90_b' and 'flux_aper90_b_avg' in http://cxc.harvard.edu/csc2/columns/fluxes.html for more details.

and (ii) the average $\log N$-$\log S$ distribution from wide-area and deep surveys (e.g., Swartz *et al.* 2011). Since the former method requires around each object the presence of blank areas wide enough to allow the reliable estimation of the interlopers, which is not always the case, we choose the $\log N$-$\log S$ method. We estimate the number of interlopers in the ULX regime in a given galaxy, by rescaling the $\log N$-$\log S$ for the *Chandra*-covered fraction of the D_{25} area of the galaxy, and integrating it down to the flux corresponding to 10^{39} erg s^{-1} for its distance. We use the $\log N$-$\log S$ from the *Chandra* Multiwavelength Project (*ChaMP*; Kim *et al.* 2007, model 'Bc' in their table 3). We account for uncertainties in the galaxy distances and the *ChaMP* $\log N$-$\log S$ parameters by Monte Carlo sampling from the corresponding Gaussian error distributions, assuming parameter independence. The expected f/b contamination in each galaxy is given in column (14) of Table 3.2.1.

The second step is to estimate the number of bona-fide ULXs in each galaxy given the number of observed sources and the previously calculated background contamination. We model the total number of sources as a mix of ULXs and interlopers, assuming both populations are Poisson distributed. We determine the posterior distribution of the number of ULXs, following the Bayesian method described in Park *et al.* (2006) with the modification that the background 'counts' in our case are not directly measured but estimated from the $\log N$-$\log S$. Specifically,

$$\begin{aligned} N_{\rm obs} &= N_{\rm ulx} + N_{\rm f/b} \\ N_{\rm ulx} &\sim {\rm Pois}(\lambda) \\ N_{\rm f/b} &\sim {\rm Pois}(\beta), \end{aligned} \tag{3.1}$$

where $N_{\rm obs}$ is the observed number of sources in each galaxy: the sum of $N_{\rm ulx}$ ULXs and $N_{\rm f/b}$ interlopers. The latter follow Poisson distributions of means λ and β, respectively, which are independent because the ULX sources in the target galaxy and the foreground/background sources are disconnected populations:

$$N_{\rm obs} \sim {\rm Pois}(\lambda + \beta) \quad \text{and} \quad P(\lambda, \beta) = P(\lambda)P(\beta). \tag{3.2}$$

To estimate the expected number of ULXs for each galaxy we compute the posterior

distribution, marginalised over β:

$$P(\lambda|N_{\rm obs}) = \int_0^\infty P(\lambda, \beta|N_{\rm obs})\, {\rm d}\beta, \tag{3.3}$$

where

$$P(\lambda, \beta|N_{\rm obs}) \propto P(N_{\rm obs}|\lambda, \beta)P(\lambda, \beta) = P(N_{\rm obs}|\lambda, \beta)P(\lambda)P(\beta). \tag{3.4}$$

In order to account for uncertainties in the parameters of the $\log N$-$\log S$ distribution, the number of interlopers β is not fixed, but allowed to vary. Specifically, the prior for β is obtained by evaluating the $\log N$-$\log S$ for varying values of its best-fitting parameters. This is performed by taking M=10000 samples of the parameters from the corresponding Gaussian distributions (best-fiting values as means, and uncertainties as standard deviations), ultimately giving M samples, β_i, which represent the distribution of β. By design, the samples β_i have equal probability to be sampled, $P(\beta_i)\propto 1$. For a uniform prior for λ, $P(\lambda)\propto 1$, and sufficiently large M, the marginalised posterior takes the form

$$P(\lambda|N_{\rm obs}) = P(N_{\rm obs}|\lambda, \beta)P(\beta)P(\lambda) \propto \sum_{i=1}^{M} P(N_{\rm obs}|\lambda, \beta_i), \tag{3.5}$$

where

$$P(N_{\rm obs}|\lambda, \beta_i) \propto (\lambda + \beta_i)^{N_{\rm obs}}\, e^{-\lambda-\beta_i}, \quad \lambda \geq 0. \tag{3.6}$$

From the resulting posterior distribution for each galaxy, we compute the mode and the highest posterior density interval corresponding to 68% probability. The observed number of sources ($N_{\rm obs}$), and the estimate on the number of ULXs ($N_{\rm ulx}$) for each galaxy in our sample are listed in Table 3.2.1 (columns (13) and (15) respectively).

To evaluate the accuracy of this method, we compare against the previously published ULX catalogue of Wang *et al.* (2016) which uses the 'blank fields' approach. We perform this comparison for the 343 galaxies in the sample of Wang *et al.* (2016) that are common with our sample. We adopt the same luminosity cut (2×10^{39} erg s^{-1}),

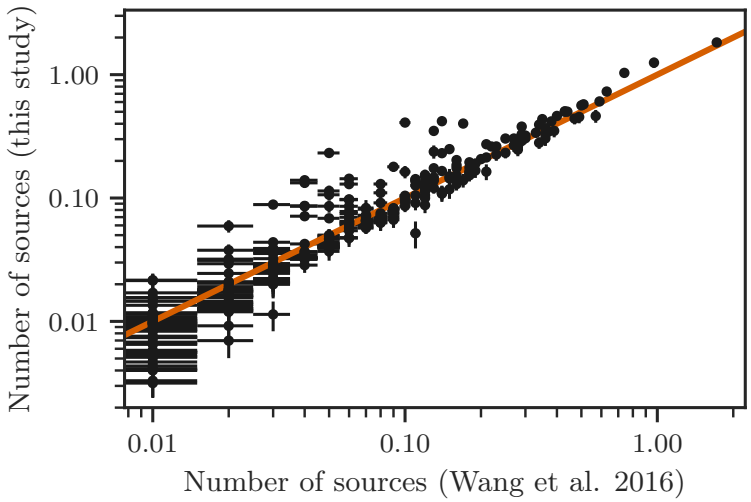


Figure 3.3.3: Comparison between the expected foreground/background number of sources from 'blank fields' approach (Wang *et al.* 2016), and log N-log S (this study), down to a luminosity limit of $2{\times}10^{39}$ erg s^{-1}. The orange line indicates the 1:1 relation. The x-error bars are equal to 0.005 counts because of the 0.01 precision of the reported values from Wang *et al.* (2016) and the y-error bars reflect the uncertainty from the log N-log S parameters.

distance and area on the sky for each galaxy as Wang *et al.* (2016). We exclude four very local galaxies from this comparison because the 2×10^{39} erg s^{-1} limit corresponds to brighter fluxes than those used to derive the *ChaMP* log N-log S. The results of the comparison of the two methods are shown in Figure 3.3.3. We find that both approaches agree in the total number of interlopers: 33.1±0.1 ('blank fields') and 36.4±0.3 (log N-log S) interlopers in the galaxies used for this comparison. The results on individual galaxies are in good agreement for the majority of them: 3σ consistency for 90% of the objects. A possible explanation for the disagreement of the two methods for the remaining 10% of the galaxies is the fact that the 'blank fields' method is based on observations of the individual galaxies, and therefore able to account for the cosmic variance at their location. However, this estimate suffers from incompleteness, due to the degradation of the PSF at the larger off-axis angles from which the background sources are sampled (note that the log N-log S estimate is on average 10% higher than the 'blank fields' estimate).

3.3.7 AGN in the host galaxies

The presence of AGN in the host galaxies can affect our investigation of ULX populations in two ways:

(i) While AGN typically have $L_X > 10^{42}$ erg s^{-1} (e.g., Brandt & Alexander 2015), they may exhibit X-ray luminosities as low as 10^{39}-10^{40} erg s^{-1} (e.g., Ho *et al.* 2001; Ghosh *et al.* 2008; Eracleous *et al.* 2002), and therefore may contaminate the sample of luminous X-ray binaries. We account for this by excluding from the demographics (but not the provided catalogues), the nuclear sources in any galaxies classified as AGN, as well as, in galaxies for which we do not have any information on their nuclear activity. We consider sources as nuclear if they are located in the central 3 arcsec region*. These sources are indicated in column (10) in Table 3.3.1. Note that this practice unavoidably removes circum-nuclear XRBs (e.g., Wang *et al.* 2016; Gong *et al.* 2016).

(ii) The IR component of the AGN emission will overestimate the inferred SFR and $M_\star$ of the host galaxy, and it will bias the measured metallicity (e.g., Mullaney *et al.* 2011; Delvecchio *et al.* 2020). While the magnitude of this effect can be small in the case of low-luminosity AGN, we take the conservative measure of excluding any AGN hosts (or galaxies with no nuclear activity information) from our scaling relations. However, scaling relations considering all galaxies (regardless of nuclear activity), labelled as 'full' sample to avoid confusion with the non-AGN sample, are presented in Appendix 3.B and are discussed in the main text when relevant.

To characterise the nuclear activity of the galaxies in our sample, we adopt the classification from the *HECATE*, which uses two sources of information to identify AGN:

(i) Stampoulis *et al.* (2019) who classified galaxies as AGN based on their location in 4- or 3-dimensional optical emission-line ratio diagnostic diagrams, using spectroscopic data from the *MPA-JHU catalogue*.

(ii) She *et al.* (2017) who investigated galaxies at $D < 50$ Mpc, observed by *Chandra*: the nuclear classifications are either adopted from the literature or determined using archival optical line-ratio spectral data.

* i.e., three times the quadratic sum of the typical positional uncertainty in the *HECATE* and the *CSC 2.0* (1 arcsec). The positional uncertainties of the sources are considered negligible since 98% of the circum-nuclear sources have positional uncertainties <3″.

A galaxy is classified as AGN host if it is identified as such in either of the two. These studies provide nuclear activity diagnostics for 539 (84%) galaxies out of the 644 host galaxies within 40 Mpc. Note, that the exclusion of AGN in the scaling relations, affects the sample of ETGs more strongly since they have higher chance of having been observed due to their nuclear activity, while spiral and irregular galaxies are usually selected for their XRB populations. This is illustrated in Figure 3.3.4 where we plot the fraction of galaxies with AGN as a function of the morphological type, in the parent galaxy sample and the galaxies with *Chandra* observations.

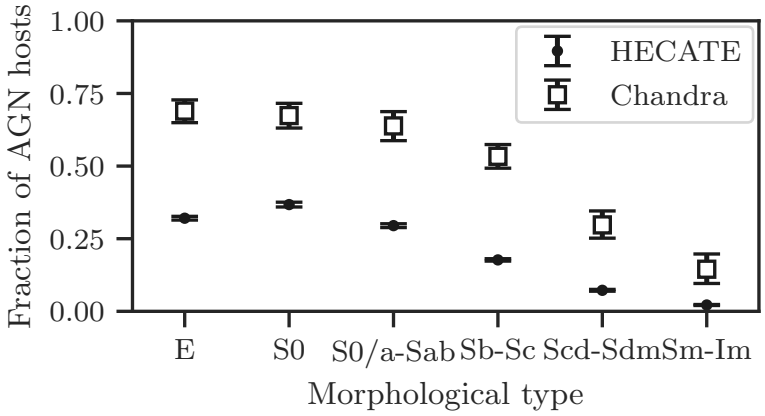


Figure 3.3.4: The AGN fraction as a function of the morphological type for the *HECATE* galaxies (points) and those hosting *Chandra* sources (squares). The error bars correspond to the 68% CI of the fractions, after accounting for uncertainties on the morphological types, and the Poisson distribution of the number of galaxies.The AGN fraction in *Chandra* targets is higher than in the general population, especially in elliptical galaxies.

3.4 Results

In the total volume of the *HECATE* we find 23043 X-ray sources, out of which 19260 are characterised as 'reliable'. In the D<40 Mpc sample which is used for the population statistics presented below, there are 16758 'reliable' X-ray sources, out of which 793 exceed the ULX limit. Of those 793 sources with L_X>10^{39} erg s^{-1} in the D<40 Mpc volume, 164 (21%) are found close to the centres of galaxies which are classified as AGN hosts, and therefore are not considered as ULX candidates in this study. This leaves a sample of 629 ULX candidates in 309 galaxies, out of which 20% are expected to be foreground/background contaminants (see §3.4.1).

3.4.1 Luminosity distribution of X-ray sources

The luminosity distribution of ULXs is crucial for probing the high-end of the luminosity function (LF) of stellar X-ray sources. The calculation of the LF of the XRBs will be presented in a separate paper. Here, we discuss the distribution of X-ray luminosities above the ULX limit.

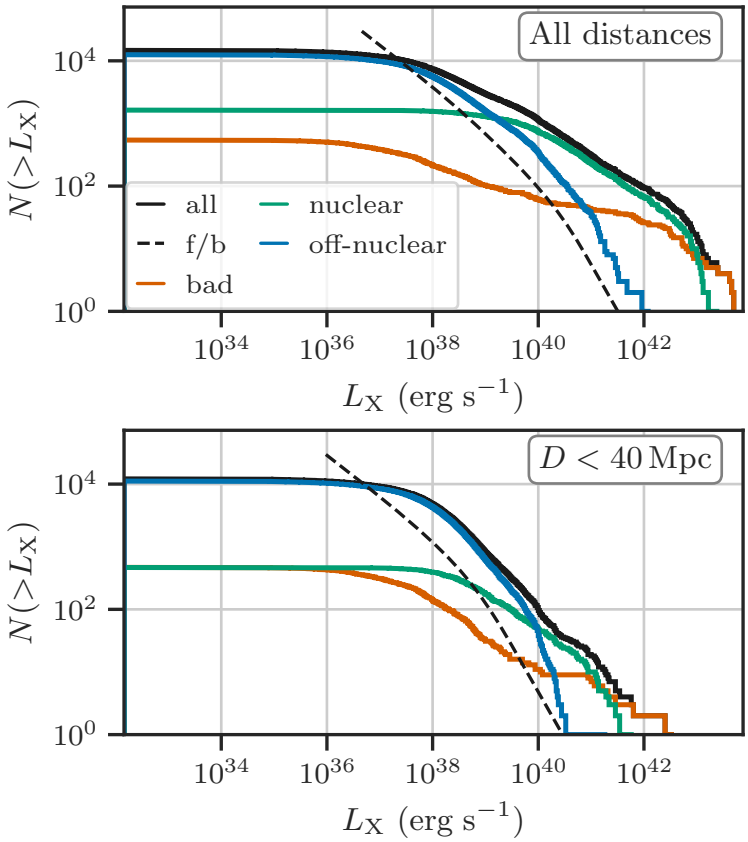


Figure 3.4.1: Cumulative number of sources, $N(>L_X)$, as a function of X-ray luminosity (L_X), for various categories: all sources (black), 'unreliable' sources (orange), off-nuclear 'reliable' sources (blue), nuclear 'reliable' sources (green). The expected number of interlopers is shown by the black dashed lines. **Top**: for all sources in our sample. **Bottom**: for sources in galaxies with D<40 Mpc.

Figure 3.4.1 shows the cumulative distribution of the luminosities of the X-ray sources in our sample in all galaxies (top panel) and in those with D<40 Mpc (bottom panel). We provide the distribution of all (black), nuclear (green), off-nuclear (blue) and 'unreliable' sources (orange; see §3.3.1). For reference, we also plot the expected

distribution of luminosities of interlopers (f/b; see §3.3.6). Since we are interested in the contamination of these interlopers within the ULX population, we convert the log N-log S distribution of the foreground/background sources to the luminosity distribution for the corresponding galaxies using their respective distances. We find that for galaxies within 40 Mpc, the background sources (dashed black line) account for ∼20% of all the sources with $L_X > 10^{39}$ erg s^{-1}. The contamination dominates the population of the off-nuclear sources at $L_X \gtrsim 10^{40.5}$ erg s^{-1}.

The gradual flattening of the luminosity distribution for the total volume (top panel in Figure 3.4.1) with decreasing luminosity is a tell-tale sign of incompleteness effects. In the case of the $D \leq 40$ Mpc distribution (bottom panel), the flattening in the distribution due to incompleteness occurs at $\sim 3 \times 10^{38}$ erg s^{-1}, below the luminosity limit for ULX candidates in this study. In addition, the typical limiting luminosity of sources detected in the galaxies (using the least luminous source) in the total volume of the *HECATE*, $\sim 4 \times 10^{39}$ erg s^{-1}, is above the ULX limit, while in galaxies with $D < 40$ Mpc, it is $\sim 1.5 \times 10^{38}$ erg s^{-1}, well below the luminosity limit for ULX candidates in this study. Therefore, our local sample of ULXs is expected to be complete.

From the top panel of Figure 3.4.1 we can see that nuclear sources outnumber the off-nuclear sources above 2×10^{39} erg s^{-1} in the full-volume sample. This is partly the result of the larger distances of galaxies in the full volume survey leading to more significant source confusion: in the dense stellar environment of the galactic cores the sources are blended, ultimately flattening the luminosity distribution. Instead, at the $D < 40$ Mpc sample, the source confusion is significantly reduced: the nuclear sources dominate the sample at a higher luminosity $\sim 10^{40}$ erg s^{-1}, as it is expected by the population of AGN.

3.4.2 Morphology of ULX hosts and spatial distribution of sources

Early ULX population studies, showed that ULXs preferentially occur in late-type galaxies, with only a small fraction ($\lesssim$20%) of elliptical galaxies hosting ULXs (e.g., Swartz *et al.* 2004; Liu *et al.* 2006; Swartz *et al.* 2011), in contrast to the recent studies of Wang *et al.* (2016) and Earnshaw *et al.* (2019). To test this, we quantify the probability for a galaxy to host ULXs as a function of its morphological type. Figure 3.4.2 shows the fraction of galaxies that host off-nuclear sources above different

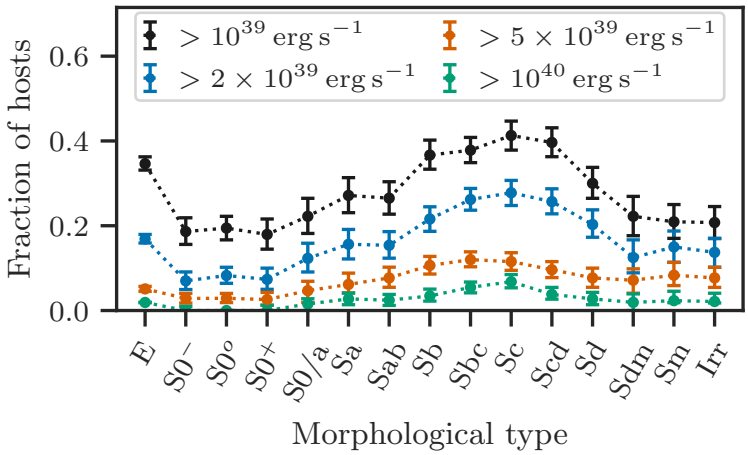


Figure 3.4.2: Fraction of galaxies in the *HECATE/CSC 2.0* hosting at least one off-nuclear source above the ULX limit (black) as a function of the morphological type. Different colours indicate different limits: $2{\times}10^{39}$ erg s^{-1} (blue), $5{\times}10^{39}$ erg s^{-1} (orange) and 10^{40} erg s^{-1} (green). The error bars indicate the 68% CIs, after accounting for the uncertainties on the morphological classifications.

luminosity thresholds, with respect to all galaxies of the same morphological type with *Chandra* observations. Sources with luminosities above 10^{39} erg s^{-1} appear to be present in about 30% of galaxies in all morphological types. There is slightly higher incidence of ULXs (~40%) in elliptical galaxies and Sb-Scd spiral galaxies, while lenticular and irregular galaxies are less likely to host ULXs (~20%). However, galaxies containing sources with $L_X{>}5{\times}10^{39}$ erg s^{-1} are typically of late type.

The spatial distribution of X-ray sources can provide insights into their nature. We would expect the surface density of ULXs to follow the distribution of starlight in the host galaxies. However, Swartz *et al.* (2011) and Wang *et al.* (2016) find a flattening of the surface density of ULXs in spiral galaxies at large galactocentric distance, in contrast to their exponential surface brightness profile. In addition, Wang *et al.* (2016) observe an excess of ULXs at large galactocentric distances in elliptical galaxies.

In order to test these observations, we quantify the spatial distribution of ULXs, by computing their surface density on the basis of their galactocentric distances, c, for off-nuclear sources with luminosities above 1, 2 and $5{\times}10^{39}$ erg s^{-1}. Using the method described in §3.3.6, we correct for the expected f/b contamination. We perform this exercise for galaxies of different morphological types and distances up to 40 Mpc. The distributions are shown in Figure 3.4.3. We do not report the number of ULXs at $c{<}0.1$

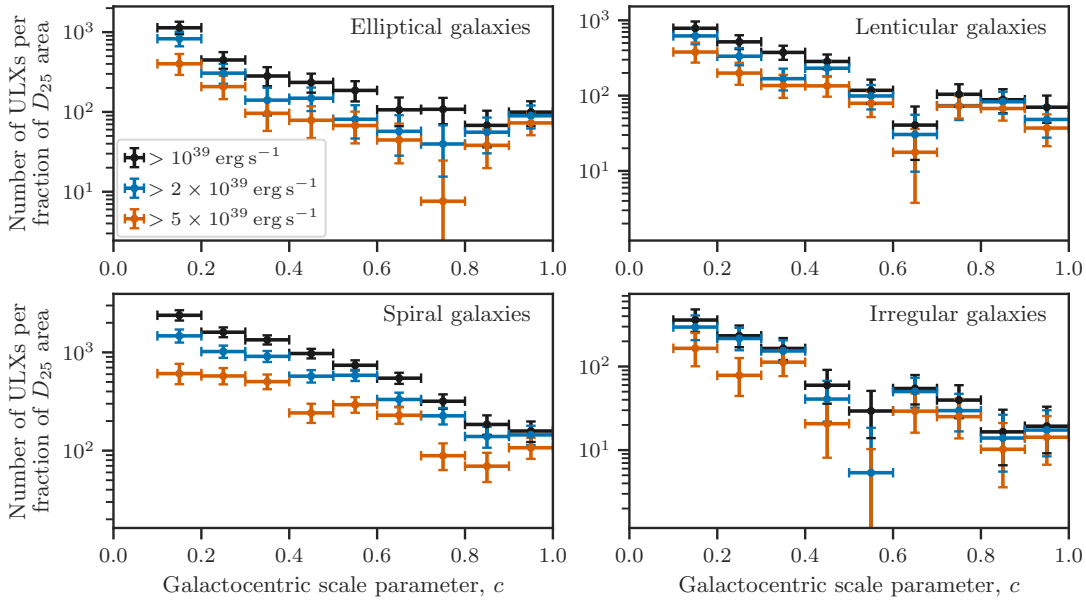


Figure 3.4.3: The density of ULXs per fraction of D_{25} area enclosed in annuli with galactocentric distances $c - dc$ to $c + dc$, as function of c, for different luminosity thresholds (colours), and morphological classes of galaxies. The estimates are corrected for f/b contamination. The dips observed in lenticular (at c~0.65) and irregular galaxies (at c~0.55) are statistical fluctuations due to the small numbers of sources; they were found to be sensitive to the binning scheme.

since it is biased by the exclusion of nuclear sources*.

We find that the surface density of ULXs follows the expected exponential trend in spiral galaxies, in contrast to Swartz *et al.* (2004) and Wang *et al.* (2016). This disagreement may be caused by foreground/background contamination since it is not accounted for in the spatial distribution analysis of Swartz *et al.* (2004) and Wang *et al.* (2016). The number of interlopers scales with area and therefore adds a constant in the surface density profile, effectively flattening the distributions. In our sample, the surface density of ULXs in elliptical galaxies flattens at large galactocentric distances as observed by Wang *et al.* (2016).

* At 40 Mpc the typical semi-major axis of galaxies in our sample is 30 arcsec, while in this study we consider nuclear regions of 3 arcsec.

3.4.3 Rate of ULXs

To quantify the number of ULXs per galaxy as a function of their luminosity for various morphological types, we consider five luminosity thresholds: 1, 2, 3, 5 and 10×10^{39} erg s^{-1}. We compute the background-corrected number of ULXs above each luminosity threshold, and its 68% confidence intervals, by accumulating the number of observed sources and expected interlopers, and applying the method described in §3.3.6. The calculation is performed for each morphological class, as well as, the total galaxy sample. Galaxies with uncertain morphological classification are excluded from this analysis (uncertainty in numerical morphological code >1; see Table 3.3.2). By considering non-AGN host galaxies, we also calculate the number of ULXs per total SFR for LTGs, or per total $M_\star$ for ETGs. In addition, we also report the number of ULXs per total $M_\star$ for all ETGs ('full'). The results are presented in Table 3.4.1.

We find that the number of ULXs per galaxy for the total population is 0.88±0.05 for $L_X>10^{39}$ erg s^{-1} and 0.20±0.02 for $L_X>5\times10^{39}$ erg s^{-1}. Comparable frequencies are found in early spirals (S0/a-Sb), while in late spirals (Sbc-Sd) they are ~1.5 times higher. In elliptical galaxies (E) the ULX frequency per galaxy is slightly lower than that of the total population. Lenticular (S0) galaxies present the lowest frequencies in all luminosity limits, with irregular galaxies following.

The number of ULXs per SFR in irregular galaxies (Sdm-Im) is higher than in spirals, in contrast to their small numbers per galaxy. Early spirals (S0/a-Sb) exhibit the lowest numbers of ULXs per SFR.

Additionally, the 'full' sample of ETGs presents lower specific ULX frequencies than the non-AGN sample by a factor of ~2. Interestingly, we find that the number of ULXs per $M_\star$ is higher in elliptical (E) than in lenticular galaxies (S0) by a factor of ~2, a result also observed by Wang *et al.* (2016). However, this trend disappears when considering the 'full' ETG sample: the specific ULX frequencies are consistent within the uncertainties.

In addition to the above analysis, we perform fits of the number of ULXs against the $M_\star$ or SFR of the galaxies. For $N_{\rm obs}$ sources with $L_X>10^{39}$ erg s^{-1} and $N_{\rm f/b}$ expected number of interlopers from the *ChaMP* log N-log S, we fit the model

$$N_{\rm obs} = \mathrm{Pois}\left(b \times M_\star + N_{\rm f/b}\right) \tag{3.7}$$

Table 3.4.1: Background-corrected number of ULXs above certain luminosity limits (first line) for different host morphological classes (first column). Square brackets indicate the total number of observed sources followed by the expected number of interlopers. The confidence intervals (68%) account for Poisson uncertainties in both the number of observed sources and the number of interlopers. **(A)**: Number of ULXs divided by the number of host galaxies (second column). The last row (total) refers to all morphological classes. **(B)**: The number of ULXs per $M_\odot\,yr^{-1}$ of SFR (third column). The last row (LTGs) refers to all late-type galaxies. **(C)**: The number of ULXs per $10^{12}\,M_\odot$ stellar mass (third column). The last row (ETGs) refers to all early-type galaxies. **(D)**: As (C), but now considering the full sample of ETGs (not excluding AGN hosts).

(A) *Number of ULXs per galaxy*

Morph. type	N_{gal}	$>10^{39}$ erg s^{-1}	$>2\times10^{39}$ erg s^{-1}	$>3\times10^{39}$ erg s^{-1}	$>5\times10^{39}$ erg s^{-1}	$>10^{40}$ erg s^{-1}
E	101	$0.68^{+0.10}_{-0.10}$ [105|36.6]	$0.32^{+0.07}_{-0.06}$ [47|15.0]	$0.23^{+0.06}_{-0.05}$ [32|8.4]	$0.16^{+0.05}_{-0.04}$ [20|3.9]	$0.03^{+0.02}_{-0.02}$ [4|1.4]
S0	99	$0.28^{+0.07}_{-0.07}$ [47|18.8]	$0.11^{+0.05}_{-0.04}$ [18|7.6]	$0.09^{+0.04}_{-0.03}$ [13|4.2]	$0.05^{+0.03}_{-0.02}$ [7|2.0]	$0.00^{+0.01}_{-0.00}$ [0|0.7]
S0/a-Sb	89	$1.02^{+0.12}_{-0.12}$ [115|24.3]	$0.48^{+0.08}_{-0.08}$ [52|9.5]	$0.38^{+0.07}_{-0.07}$ [39|5.2]	$0.25^{+0.06}_{-0.05}$ [25|2.5]	$0.09^{+0.04}_{-0.03}$ [9|0.9]
Sbc-Sd	166	$1.38^{+0.10}_{-0.09}$ [258|28.5]	$0.76^{+0.07}_{-0.07}$ [138|11.7]	$0.48^{+0.06}_{-0.05}$ [87|6.6]	$0.28^{+0.04}_{-0.04}$ [50|3.1]	$0.12^{+0.03}_{-0.03}$ [21|1.1]
Sdm-Im	37	$0.48^{+0.13}_{-0.11}$ [20|2.1]	$0.30^{+0.10}_{-0.08}$ [12|0.8]	$0.26^{+0.10}_{-0.08}$ [10|0.4]	$0.18^{+0.08}_{-0.06}$ [7|0.2]	$0.03^{+0.04}_{-0.02}$ [1|0.1]
Total	492	$0.88^{+0.05}_{-0.05}$ [545|110.3]	$0.45^{+0.03}_{-0.03}$ [267|44.5]	$0.32^{+0.03}_{-0.03}$ [181|24.9]	$0.20^{+0.02}_{-0.02}$ [109|11.7]	$0.06^{+0.01}_{-0.01}$ [35|4.2]

(B) *Number of ULXs per* $M_\odot\,yr^{-1}$ *SFR in non-AGN, late-type galaxies*

Morph. type	N_{gal}	SFR	$>10^{39}$ erg s^{-1}	$>2\times10^{39}$ erg s^{-1}	$>3\times10^{39}$ erg s^{-1}	$>5\times10^{39}$ erg s^{-1}	$>10^{40}$ erg s^{-1}
S0/a-Sb	23	86.5	$0.29^{+0.07}_{-0.06}$ [30|4.6]	$0.17^{+0.05}_{-0.04}$ [16|1.7]	$0.14^{+0.05}_{-0.04}$ [13|1.0]	$0.06^{+0.03}_{-0.02}$ [6|0.4]	$0.02^{+0.02}_{-0.01}$ [2|0.2]
Sbc-Sd	76	97.9	$0.78^{+0.10}_{-0.09}$ [87|10.6]	$0.40^{+0.07}_{-0.06}$ [43|4.1]	$0.22^{+0.05}_{-0.05}$ [24|2.3]	$0.12^{+0.04}_{-0.03}$ [13|1.1]	$0.07^{+0.03}_{-0.02}$ [7|0.4]
Sdm-Im	20	2.3	$2.39^{+1.21}_{-0.91}$ [6|0.4]	$1.22^{+0.91}_{-0.61}$ [3|0.1]	$1.25^{+0.91}_{-0.61}$ [3|0.1]	$0.84^{+0.79}_{-0.48}$ [2|0.0]	$0.00^{+0.48}_{-0.00}$ [0|0.0]
LTGs	119	186.7	$0.58^{+0.06}_{-0.08}$ [123|15.5]	$0.30^{+0.04}_{-0.04}$ [62|6.0]	$0.20^{+0.04}_{-0.03}$ [40|3.3]	$0.10^{+0.03}_{-0.02}$ [21|1.6]	$0.05^{+0.02}_{-0.01}$ [9|0.6]

(C) *Number of ULXs per* $10^{12}\,M_\odot$ *stellar mass in non-AGN, early-type galaxies*

Morph. type	N_{gal}	$M_\star$	$>10^{39}$ erg s^{-1}	$>2\times10^{39}$ erg s^{-1}	$>3\times10^{39}$ erg s^{-1}	$>5\times10^{39}$ erg s^{-1}	$>10^{40}$ erg s^{-1}
E	22	0.82	$23.7^{+6.5}_{-5.6}$ [25|5.4]	$19.2^{+5.5}_{-4.7}$ [18|2.2]	$16.7^{+5.1}_{-4.3}$ [15|1.2]	$13.9^{+4.6}_{-3.8}$ [12|0.6]	$1.0^{+1.7}_{-0.9}$ [1|0.2]
S0	28	0.73	$10.1^{+4.8}_{-3.9}$ [10|2.6]	$8.1^{+4.1}_{-3.2}$ [7|1.0]	$6.0^{+3.5}_{-2.6}$ [5|0.6]	$3.7^{+2.9}_{-2.0}$ [3|0.3]	$0.0^{+1.5}_{-0.0}$ [0|0.1]
ETGs	50	1.56	$17.3^{+4.0}_{-3.6}$ [35|8.0]	$13.9^{+3.4}_{-3.0}$ [25|3.2]	$11.7^{+3.1}_{-2.7}$ [20|1.8]	$9.1^{+2.7}_{-2.3}$ [15|0.8]	$0.5^{+0.9}_{-0.5}$ [1|0.3]

(D) *Number of ULXs per* $10^{12}\,M_\odot$ *stellar mass in all early-type galaxies*

Morph. type	N_{gal}	$M_\star$	$>10^{39}$ erg s^{-1}	$>2\times10^{39}$ erg s^{-1}	$>3\times10^{39}$ erg s^{-1}	$>5\times10^{39}$ erg s^{-1}	$>10^{40}$ erg s^{-1}
E (full)	96	7.8	$8.72^{+1.34}_{-1.26}$ [105|36.6]	$4.09^{+0.91}_{-0.83}$ [47|15.0]	$3.01^{+0.76}_{-0.68}$ [32|8.4]	$2.05^{+0.61}_{-0.53}$ [20|3.9]	$0.33^{+0.30}_{-0.21}$ [4|1.4]
S0 (full)	98	5.2	$5.37^{+1.36}_{-1.24}$ [47|18.8]	$1.98^{+0.87}_{-0.74}$ [18|7.6]	$1.67^{+0.75}_{-0.62}$ [13|4.2]	$0.95^{+0.57}_{-0.44}$ [7|2.0]	$0.00^{+0.22}_{-0.00}$ [0|0.7]
ETGs (full)	194	13.1	$7.38^{+0.96}_{-0.91}$ [152|55.4]	$3.24^{+0.64}_{-0.59}$ [65|22.5]	$2.47^{+0.54}_{-0.49}$ [45|5.9]	$1.61^{+0.42}_{-0.37}$ [27|5.9]	$0.14^{+0.16}_{-0.12}$ [4|2.1]

Table 3.4.2: Fitting results for the scaling factor a (Equation 3.8) for all LTGs and different morphological classes, and the scaling factor b (Equation 3.7) for all ETGs, elliptical (E) and lenticular (S0) galaxies. These results are plotted in Figure 3.4.4.

Morphology	$N_{\rm gal}$	$N_{\rm src}$	$N_{\rm f/b}$	a (ULXs per $M_\odot$ yr^{-1})
LTGs	119	123	17.0	$0.51^{+0.06}_{-0.06}$
S0/a-Sab	11	18	2.9	$0.23^{+0.07}_{-0.07}$
Sb	12	12	2.2	$0.34^{+0.14}_{-0.12}$
Sbc	14	26	4.8	$0.44^{+0.12}_{-0.11}$
Sc	17	19	2.6	$0.51^{+0.15}_{-0.13}$
Scd	22	23	2.4	$1.54^{+0.39}_{-0.34}$
Sd-Sdm	23	19	1.7	$2.29^{+0.67}_{-0.57}$
Sm-Im	20	6	0.5	$2.16^{+1.19}_{-0.90}$
Morphology	$N_{\rm gal}$	$N_{\rm src}$	$N_{\rm f/b}$	b (ULXs per $10^{12}M_\odot$)
ETGs	50	35	8.3	$15.1^{+3.9}_{-3.6}$
E	22	25	5.5	$21.9^{+6.4}_{-5.7}$
S0	28	10	2.8	$8.7^{+4.7}_{-3.8}$
ETGs (full)	195	152	57.6	$6.3^{+1.0}_{-0.9}$
E (full)	96	104	37.5	$7.5^{+1.3}_{-1.3}$
S0 (full)	99	47	20.0	$4.8^{+1.4}_{-1.2}$

for all ETGs with robust morphological classifications (see caption of Table 3.3.2), and the subdivisions of elliptical and lenticular galaxies, using the Maximum Likelihood method. In a similar fashion, we fit the model

$$N_{\rm obs} = \mathrm{Pois}\left(a \times \mathrm{SFR} + N_{\rm f/b}\right) \tag{3.8}$$

for the LTGs and five sub-populations: S0/a-Sab, Sb, Sbc, Sc, Scd, Sd-Sdm and Sm-Im*. The results are listed in Table 3.4.2.

For the scaling with $M_\star$ in ETGs we find b=$15.1^{+3.9}_{-3.6}$ ULXs per $10^{12}M_\odot$, while in elliptical galaxies it is significantly higher ($21.9^{+6.4}_{-5.7}$ ULXs per $10^{12}M_\odot$) than in lenticular galaxies ($8.7^{+4.7}_{-3.8}$ ULXs per $10^{12}M_\odot$). However, in the full ETG sample (i.e. including AGN hosts), the specific ULX frequencies are lower than those in the non-AGN ETGs ($6.3^{+1.0}_{-0.9}$ ULXs per $10^{12}M_\odot$). See §3.5.4 for an explanation of this difference.

* The selection of the ranges of morphological types ensured that at least ten galaxies were contributing to the statistical estimates.

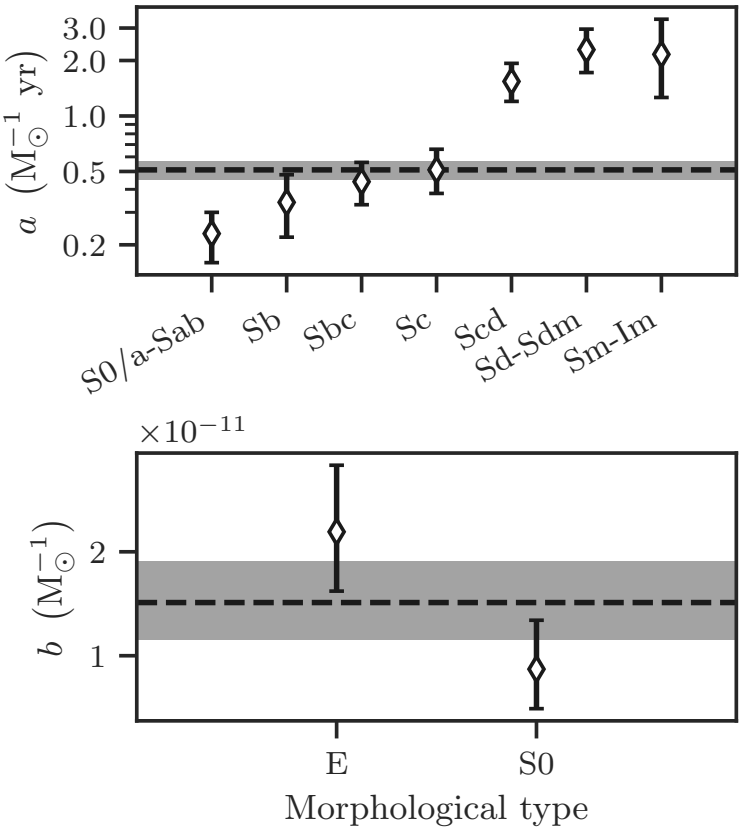


Figure 3.4.4: Fitting results for the SFR and $M_\star$ scaling factors for different morphological classes. **Top**: the scaling parameter a (Equation 3.8) for late-type galaxies (line) and 68% CI (grey band), and for various late-type morphological classes (black error bars). **Bottom**: same as the top panel, but for the scaling parameter b (Equation 3.7) for early-type (line and band), elliptical and lenticular galaxies (error bars).

In LTGs, we find that the scaling with SFR, a, is 0.51 ± 0.06 ULXs per $M_\odot\,yr^{-1}$ (horizontal line in the top panel of Figure 3.4.4) and that it monotonically increases with morphological type: from 0.23 ULXs per $M_\odot\,yr^{-1}$ (S0/a-Sab) to 2.16 ULXs per $M_\odot\,yr^{-1}$ (Sm-Im).

3.4.4 SFR and stellar mass scaling in late-type galaxies

In order to account for the contribution of ULXs associated with LMXBs*, we perform a joint fit of the number of ULXs in LTGs with respect to both their SFR and

* E.g., GRS1915+105-type systems; Greiner *et al.* (2001)

$M_\star$.

The correlation can be visualised by binning the galaxies by SFR and $M_\star$ and computing the average number of ULXs per galaxy, for each bin, after removing the f/b contamination (see §3.3.6). We plot the result in Figure 3.4.5, where we see a trend for galaxies with higher SFR and $M_\star$ to host larger numbers of ULXs. This trend becomes stronger in regions of high sSFR (indicated by the diagonal lines).

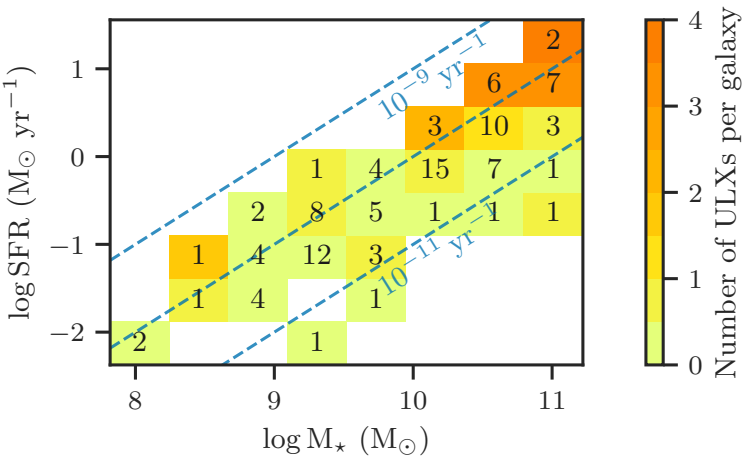


Figure 3.4.5: The mean, background-corrected number of ULXs ($N_{\rm ulx}$) per galaxy (color scale) as a function of SFR and $M_\star$, in non-AGN LTGs. The diagonal dashed lines correspond to indicative specific SFRs, while the numbers in boxes denote the number of galaxies in each bin. We see a trend for more ULXs in galaxies with high SFR, $M_\star$, and sSFR.

While SFR and $M_\star$ are known to be correlated in star-forming galaxies (e.g., Rodighiero *et al.* 2011; Speagle *et al.* 2014; Maragkoudakis *et al.* 2017), the SFR is expected to be the primary parameter correlated with the population of ULXs. To study the dependence of $N_{\rm ulx}$ on both parameters, we fit the model

$$N_{\rm obs} \sim {\rm Pois}\left(\alpha \times {\rm SFR} + \beta \times M_\star + N_{\rm f/b}\right) \tag{3.9}$$

where $N_{\rm obs}$ is the total number of observed sources with $L_{\rm X}{>}10^{39}$ erg s^{-1}, α and β are the scaling factors that will be fitted, and $N_{\rm f/b}$ is the expected number of interlopers (computed in §3.3.6). The model is applied to all LTGs with robust morphological classifications (see Tables 3.2.1, 3.3.2) and the two sub-populations of early spirals (S0/a-Sbc), and late spirals / irregular galaxies (Sc-Im). The results are listed in Table 3.4.3,

Table 3.4.3: Mode and 68% Highest Posterior (marginalised) Density Intervals of the scaling parameters α and β (see Equation 3.9) for all LTGs and their 'early' and 'late' subdivisions (see Table 3.3.2). For each fit we report the number of galaxies ($N_{\rm gal}$), sources ($N_{\rm src}$) and interlopers ($N_{\rm f/b}$). See below for the joint distributions (Figure 3.4.6).

Sample	$N_{\rm gal}$	$N_{\rm src}$	$N_{\rm f/b}$	α $(\mathrm{M_\odot\,yr^{-1}})^{-1}$	β $(10^{12}\,\mathrm{M_\odot})^{-1}$
LTGs	106	117	15.8	$0.45^{+0.06}_{-0.09}$	$3.3^{+3.8}_{-3.2}$
S0/a-Sbc	37	56	9.9	$0.16^{+0.08}_{-0.08}$	$11.2^{+5.2}_{-5.6}$
Sc-Im	69	61	5.9	$0.98^{+0.11}_{-0.20}$	<6.6

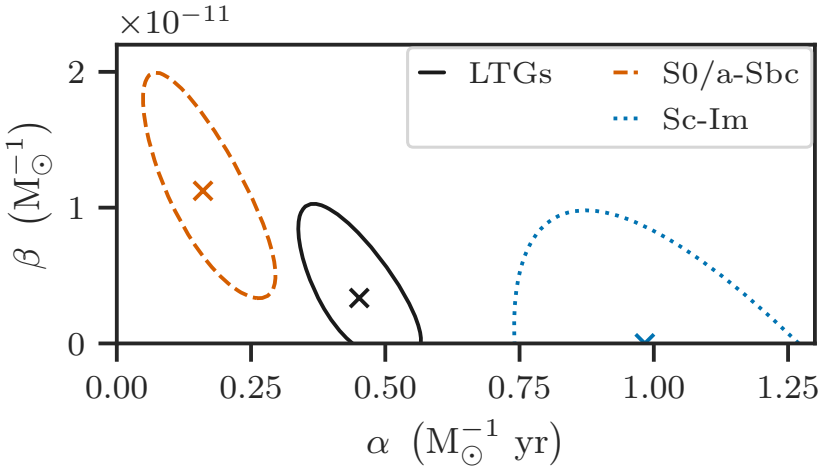


Figure 3.4.6: The best-fitting values (X symbols) and, 68% confidence regions (lines) of the scaling parameters *a* and *b* of Equation 3.9 for all LTGs (solid), and the 'early' (S0/a-Sbc; orange) and 'late' (Sc-Im; blue) spiral galaxies. See Figure 3.4.3 for marginalised results.

while the joint posterior distributions of α and β are shown in Figure 3.4.6.

The best-fitting value of the SFR scaling factor for LTGs is α=$0.45^{+0.06}_{-0.09}$ ULXs per $\mathrm{M_\odot\,yr^{-1}}$ while the $M_\star$ scaling factor is β=$3.3^{+3.8}_{-3.2}$ ULXs per $10^{12}\mathrm{M_\odot}$. For the early-type spirals we find lower α=0.16±0.08 ULXs per $\mathrm{M_\odot\,yr^{-1}}$ and higher β=$11.2^{+5.2}_{-5.6}$ ULXs per $10^{12}\mathrm{M_\odot}$, while for the late-type spirals the situation is inverted, i.e. α=$0.98^{+0.11}_{-0.20}$ ULXs per $\mathrm{M_\odot\,yr^{-1}}$ and β is consistent with zero (<6.6 ULXs per $10^{12}\mathrm{M_\odot}$).

3.5 Discussion

3.5.1 Comparison with other ULX surveys

Our estimate for the scaling of the number of ULXs with SFR (0.51 ± 0.06 ULXs per $M_\odot\ yr^{-1}$; Table 3.4.2) is four times lower than that estimated in Swartz *et al.* (2011; 2 ULXs per $M_\odot\ yr^{-1}$). This is the result of differences in: (i) the selection of host galaxy sample, and (ii) the method used in the calculation of the X-ray fluxes. When we account for these differences, we find consistent results as discussed in detail in Appendix §3.C.1.

Furthermore, we would expect our results to agree with those of Wang *et al.* (2016), since they also use *Chandra* observations for a similarly large sample of host galaxies (343 galaxies) to study the ULX content in nearby galaxies. However, Wang *et al.* (2016) consider ULXs at twice our luminosity threshold (i.e., 2×10^{39}) and at larger separations from the galaxy centres ($2{\times}D_{25}$ area instead of D_{25}). After accounting for these differences, and a small offset between the computed X-ray fluxes resulting from different methods, we find similar frequency of ULXs in all galaxies, and separately for their different morphological classes (Table 3.4.1). See Appendix §3.C.2 for details of this comparison.

Earnshaw *et al.* (2019), using a sample of 248 galaxies with sensitivity limit below the ULX limit in their X-ray samples, found that one out of three galaxies host at least one ULX, with spiral galaxies having a slightly higher fraction (~40%) than elliptical galaxies (~30%). This is in agreement with our results (see Figure 3.4.2): the fraction of ULX hosts in galaxies of different morphological types is between 20% and 40%, with the peak at Sc galaxies, and a fraction of ~35% in elliptical galaxies.

3.5.2 Dependence of number of ULXs on SFR and stellar mass in star-forming galaxies

In §3.4.3, we find the number of ULXs in LTGs to be $0.51{\pm}0.06$ per $M_\odot\ yr^{-1}$ (see Table 3.4.2), consistent with the expectation from the Mineo *et al.* (2012) HMXB-LF, of 0.56 ULXs per $M_\odot\ yr^{-1}$. We observe a dependence of the scaling factor per SFR (parameter a in Equation 3.8 and Figure 3.4.4) on the morphological type; it monotonically increases from $0.23{\pm}0.07$ to $2.16^{+1.19}_{-0.90}$ ULXs per $M_\odot\ yr^{-1}$ from S0/a-Sab to Sm-Im galaxies (Table 3.4.2). The higher scaling factor in late spiral and irregular

galaxies can be attributed to their lower metallicity with respect to early spiral galaxies (e.g., González Delgado *et al.* 2015): as discussed in §3.5.3 low metallicity galaxies show an excess of ULXs. Figure 3.5.1 shows the metallicity distribution for different morphological types in the *HECATE* and our host galaxy sample. We see that the average metallicity quickly drops for galaxies later than Sc, the same galaxies for which the scaling factor (see Table 3.4.2) increases from 0.51 to 1.54 ULXs per $M_{\odot}$ yr^{-1}. However, these trends do not account for another important factor which cannot be tested in the current sample: late-type galaxies are more prone to short and intense star-formation episodes, which might increase their ULX content significantly (e.g., Wiktorowicz *et al.* 2017), although the effect of metallicity appears to have a stronger effect in the X-ray output of a galaxy (e.g., Fragos *et al.* 2013a).

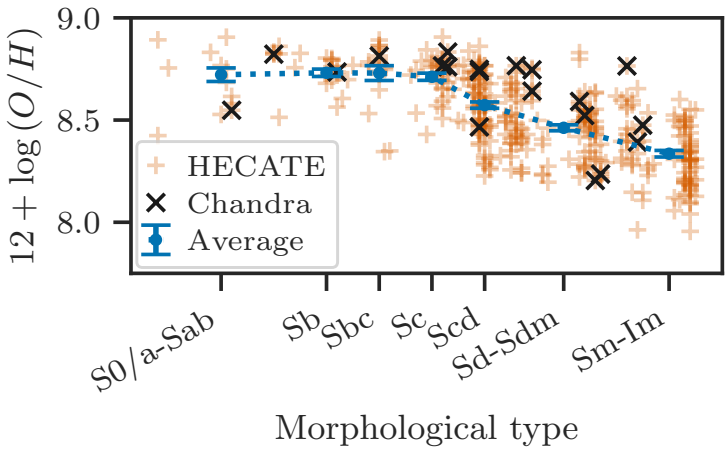


Figure 3.5.1: The metallicity, $12 + \log(O/H)$, as a function of the morphological type in the host ('Chandra'; black 'x' markers) and parent ('HECATE'; orange plus markers) galaxy samples. Note, that only non-AGN galaxies with robust morphological classification and D<40 Mpc are shown. For the morphological types that correspond to the results of the scaling of ULXs with SFR, a, in Table 3.4.2, we show the mean metallicity and its standard error (blue errorbars). Scd galaxies and later, present the lower metallicities, partly explaining the result that in the same galaxies a is significantly higher than the average in LTGs.

In order to account for the LMXB contribution in LTGs, in §3.4.3 we computed the scaling parameters α and β for the linear relation between number of observed sources with L_X>10^{39} erg s^{-1} and both SFR and $M_{\star}$ (Equation 3.9). The value of $\alpha = 0.45^{+0.06}_{-0.09}$ ULXs per $M_{\odot}$ yr^{-1} (see Figure 3.4.3) for all LTGs is somewhat smaller, but consistent with the value of $a = 0.51 \pm 0.06$ ULXs per $M_{\odot}$ yr^{-1} found using the model

of Equation 3.8 where only the SFR scaling is considered (see Table 3.4.2). The smaller scaling when accounting for the contribution of the $M_\star$ is the result of the small fraction of the ULX population that is associated with the old stellar population (and consequently the $M_\star$) in spiral galaxies. The results of the fits for early and late spirals (see Figure 3.4.3 and Figure 3.4.6) illustrate that the $M_\star$ contribution is significant in early spirals (S0/a-Sbc) at the 2σ-level, while it can be neglected in late spirals (β<6.6 ULXs per $10^{12}\mathrm{M}_\odot$ with most probable value 0.0).

Recently, Lehmer *et al.* (2019) constructed luminosity functions of XRBs as a function of both SFR and $M_\star$ to account for the contribution of both LMXBs and HMXBs. Using the best-fitting parameters for their full sample (see table 4 in Lehmer *et al.* 2019) we integrated the LF above the ULX limit ($10^{39}\,\mathrm{erg\,s^{-1}}$). We find that they predict $N_{\rm ulx} = a_{\rm L19}\mathrm{SFR} + b_{\rm L19}M_\star$, where $\alpha_{\rm L19} = 0.62\pm0.08$ ULXs per $\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ and $\beta_{\rm L19} = 18^{+23}_{-11}$ ULXs per $10^{12}\mathrm{M}_\odot$. The scaling with SFR (α) is consistent at the 1σ-level with our findings for all LTG galaxies ($0.45^{+0.06}_{-0.09}$ ULXs per $\mathrm{M}_\odot\,\mathrm{yr}^{-1}$; see Equation 3.9). The scaling with $M_\star$ (β) is highly uncertain in the ULX regime, but also consistent at the 1σ-level with the one we find for all LTG galaxies ($3.3^{+3.8}_{-3.2}$ ULXs per $10^{12}\mathrm{M}_\odot$).

Finally, the above results are consistent with the qualitative picture shown in Figure 3.4.5; the number of ULXs in LTGs increases with both SFR and $M_\star$. Note that the trend of galaxies hosting larger population of ULXs at higher sSFR (see diagonal lines), may have a trivial explanation: ULXs being primarily associated with young stellar populations, are more abundant in galaxies with higher SFR and/or lower mass. However, an age effect may be at the play: starbursts have high sSFR, by definition, and are expected to have high formation rate of BH ULXs, which dominate the population at ~5 Myr (e.g. Fragos *et al.* 2013b).

Such an age effect will manifest as an excess of ULXs in high sSFR galaxies compared to the expectation from the average SFR-$M_\star$ scaling relation based on all LTGs in our sample. We assess this by defining the *excess* of ULXs,

$$\mathrm{excess} = \log \frac{N_{\rm obs}}{N_{\rm exp}} \tag{3.10}$$

where $N_{\rm obs}$ is the number of ULX candidates, and $N_{\rm exp}$ is the expected number of sources according to the model in Equation 3.9 and its best-fitting values (Figure 3.4.3). In order to explore the possible dependence of the ULX excess on SFR and $M_\star$, we plot

in Figure 3.5.2 the ULX excess of galaxies as a function of their SFR and $M_\star$. We do not see a dependence on the sSFR; instead, it is clear that low-mass galaxies present an excess of ULXs, in agreement with Swartz *et al.* (2008). Since low-mass galaxies tend to present lower metallicities (e.g., Kewley & Ellison 2008) we interpret this excess as likely being related to the metallicity of their hosts.

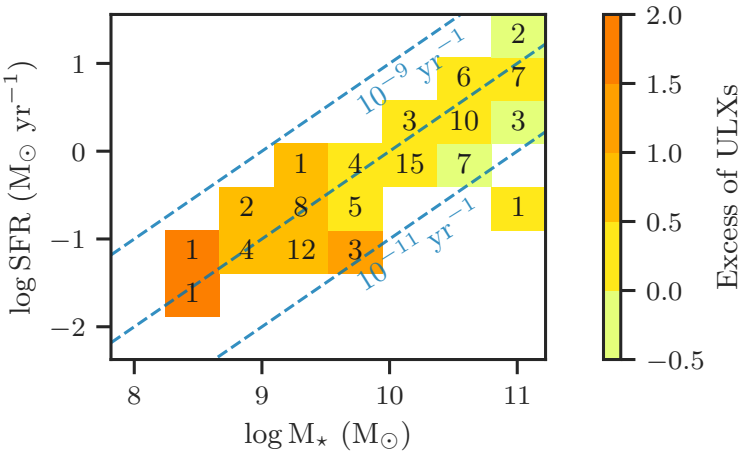


Figure 3.5.2: The mean excess of ULXs (colour scale; as expressed in Equation 3.10) in LTGs as a function of SFR and $M_\star$. The diagonal lines correspond to indicative specific SFRs. The numbers in the boxes denote the number of galaxies in each bin. Despite their small numbers, the data show that galaxies with low masses exhibit a clear excess in the number of ULXs over the expectation.

3.5.3 Excess of ULXs in low-metallicity galaxies

There is a growing observational body of evidence for an excess of ULXs in low-metallicity galaxies (e.g., Soria *et al.* 2005; Mapelli *et al.* 2010; Prestwich *et al.* 2013; Brorby *et al.* 2014; Tzanavaris *et al.* 2016). This trend can be interpreted theoretically in the context of the weaker stellar winds in low-metallicity stars. The stars retain higher fraction of their initial mass, and as a consequence, more massive BHs are formed, with smaller orbital separation due to weaker angular momentum losses (e.g., Heger *et al.* 2003; Belczynski *et al.* 2010; Marchant *et al.* 2017). In addition, the tighter orbits result in an increased fraction of HMXBs that enter a Roche-lobe overflow phase, which being a more efficient accretion mechanism than stellar winds, leads to more luminous X-ray sources (Linden *et al.* 2010).

To investigate the correlation of the ULX population with metallicity, we plot in

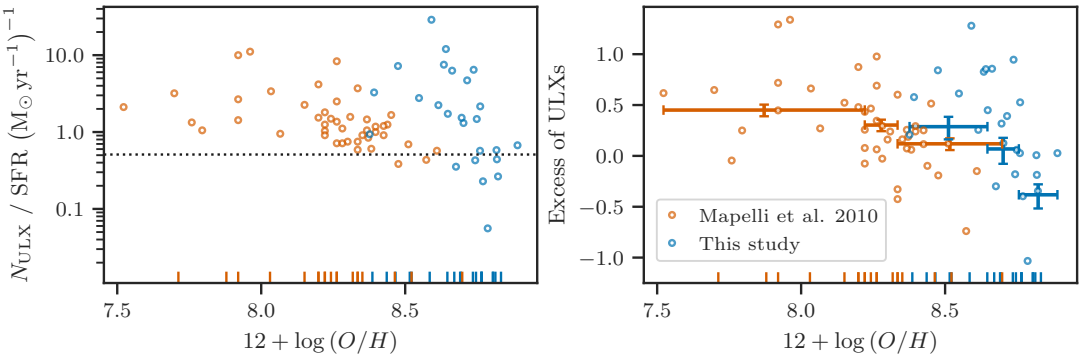


Figure 3.5.3: Left panel: the number of ULXs per SFR unit ULXs for LTGs in our sample (blue) with metallicity estimates (blue points), and that of Mapelli *et al.* (2010) (orange). The metallicities of the galaxies with no sources above the ULX limit, are shown as ticks on the x-axis. Low-metallicity galaxies present an excess of ULXs with respect to the average scaling (dotted horizontal line.) Right panel: same as the left panel but now the y-axis is the excess of ULXs as defined in Equation 3.10. We find a significant anti-correlation (Kendall rank correlation coefficient τ=−0.43 with p-value 0.002). We bin the galaxies to reduce the stochasticity of the ULX excess (see §3.5.3) and compute the median and 68% CIs of the excess as a function of the metallicity (blue error bars). We repeat the same procedure for the ULX excess in the sample of Mapelli *et al.* (2010) (orange) which also exhibits significant anti-correlation (τ=−0.40 and p=0.001).

Figure 3.5.3 (left) the number of ULXs per SFR as function of the host galaxy metallicity for the 44 galaxies with metallicity measurements in the *HECATE* (blue circles). We see an excess of ULXs in low metallicities with respect to the average relation shown with the dotted line. In the right panel of Figure 3.5.3 we plot the ULX excess (Equation 3.10) against metallicity. We see that the frequency of ULXs indeed increases with decreasing metallicity (Kendall rank correlation coefficient $\tau = -0.43$ with p-value of 0.002). For comparison, we also plot in the same figure the excess of ULXs computed from the sample in Mapelli *et al.* (2010) (orange circles) using their reported values for (i) the number of ULXs, (ii) the expected background contamination, and (iii) the SFRs of the host galaxies*. To reduce the Poisson noise, the galaxies are grouped in metallicity bins (defined to have a similar number of objects in each bin and always more than eight) shown as x-axis error bars in Figure 3.5.3. The central values and the error

* Since Mapelli *et al.* (2010) do not provide $M_\star$ estimates, which are needed to compute N_{exp} in Equation 3.10, we obtain our own estimates for N_{exp} using the SFR and the scaling constant of 0.51 ULXs per $M_\odot\,yr^{-1}$, determined from fits presented in §3.4.3 (cf. Equation 3.8). In addition, metallicities from Mapelli *et al.* (2010) were converted from solar units ($Z_\odot$) using their adopted solar metallicity $12 + \log(O/H)_\odot$ =8.92.

bar length in the y-axis correspond to the median and the 68% confidence interval of the ULX excess, computed by accounting for Poisson uncertainty of the number of sources and interlopers.

Based on the binned statistics in Figure 3.5.3, we find that the galaxies with the lowest metallicities in our sample (corresponding to 0.3-0.5 $Z_\odot$) host more ULXs per SFR by a factor of ~2, in comparison to galaxies of intermediate metallicity (0.5-0.7 $Z_\odot$) which present no excess of ULXs. Interestingly, galaxies with near-solar metallicity (>0.7 $Z_\odot$) present a deficiency of ULXs; they host half of the expected ULX population.

The same trend is observed in the sample of Mapelli *et al.* (2010). However, there seems to be a small horizontal offset of ~0.25 dex between our study and that of Mapelli *et al.* (2010). We attribute this offset to the different metallicity calibrations* which can have systematic biases up to 0.7 dex (see fig. 2 in Kewley & Ellison 2008). Using eight common galaxies in our sample and that of Mapelli *et al.* (2010), we find that the mean offset between the metallicities is 0.28 ± 0.09.

In conclusion, an excess of ULXs is linked with low-metallicities. This is also in line with our result that the ULX-SFR scaling factor is significantly higher in later-type galaxies, for which the metallicity is lower (see §3.5.2). This excess has direct implications for the XRB content of the high-redshift Universe. The mean metallicity of galaxies at z~2.5 was only ~0.1 $Z_\odot$ (e.g., Madau & Dickinson 2014). Indeed, an excess of the integrated X-ray luminosity per unit SFR is seen in observational studies of high-redshift galaxies (e.g., Lehmer *et al.* 2005, 2016; Basu-Zych *et al.* 2013a,b, 2016; Brorby *et al.* 2016; Fornasini *et al.* 2019, 2020; Svoboda *et al.* 2019). Our results indicate that this excess is the result of a larger population of luminous X-ray sources per unit SFR in lower metallicities. However, we cannot exclude the possibility that the stellar population age also plays a role on the ULX excess. Since the metallicity and age can vary by region in a galaxy, investigation on sub-galactic scales can help to disentangle their relative effects on the XRB populations (cf. Anastasopoulou *et al.* 2019; Lehmer *et al.* 2019; Kouroumpatzakis *et al.* 2020).

* In our sample we use the metallicity estimates in the *HECATE* which were calculated via the O III-N II calibration in Pettini & Pagel (2004), while the metallicities in Mapelli *et al.* (2010) are based on many different calibrations, mainly those in Pilyugin 2001 and Pilyugin & Thuan 2005.

3.5.4 ULXs and old stellar populations

ULXs in elliptical galaxies (e.g., David *et al.* 2005) are considered to belong to the high-end of the LMXB-LF (e.g., Swartz *et al.* 2004; Plotkin *et al.* 2014). Notably, rejuvenation of stellar populations due to galaxy mergers might also produce additional ULXs (Zezas *et al.* 2003; Raychaudhury *et al.* 2008; Kim & Fabbiano 2010). In addition, it is possible that a small population of ULXs are dynamically formed in GCs (e.g., Maccarone *et al.* 2007; Dage *et al.* 2020). Indeed, we find evidence that a small but significant population of ULXs in elliptical galaxies resides at large galactocentric distances (see §3.4.2), i.e., not following the $M_\star$ distributions. Literature review of the hosts of these sources showed evidence for recent merger activity, or large GC populations, indicating that these ULXs could be associated with GCs, given the flatter distributions of GCs and their LMXB populations with respect to the stellar light (e.g., Kim *et al.* 2006).

In §3.4.3, based on the fit of the number of ULXs against the $M_\star$ of ETGs (see Equation 3.7), we find $15.1^{+3.9}_{-3.6}$ ULXs per $10^{12}\mathrm{M}_\odot$ in the non-AGN sample. However, it is higher by a factor of 2-3 than the expectation from the LMXB-LF of Zhang *et al.* (2012) (5 ± 2 ULXs per $10^{12}\mathrm{M}_\odot$), and the specific ULX frequency in Plotkin *et al.* (2014) (6.2 ± 1.3 ULXs per $10^{12}\mathrm{M}_\odot$) and Walton *et al.* (2011) ($\sim$7 ULXs per $10^{12}\mathrm{M}_\odot$). While these studies address possible contamination from AGN in their X-ray source samples, they still consider (except for Zhang *et al.* 2012) the total K-band luminosity of the galaxies as a tracer of the $M_\star$ even if the galaxy hosts an AGN. The contamination by the AGN would lead to an overestimation of the $M_\star$, and consequently an underestimation of the specific ULX frequency. To quantify this effect, in Appendix 3.B we compute the specific ULX frequency in the full sample (including galaxies hosting AGN, but still excluding their nuclear sources). We find $6.3^{+1.0}_{-0.9}$ ULXs per $10^{12}\mathrm{M}_\odot$ in good agreement with the literature estimates.

Why do non-AGN ETGs exhibit higher specific ULX frequency than the 'full' ETG sample? As we show in Appendix 3.B the presence of an AGN does not significantly affect the observed K-band luminosity and therefore the measured $M_\star$. The AGN contribution is <10% of the total K-band luminosity (Bonfini et al. 2020, submitted). However, the full sample extends to much larger masses than the non-AGN ETG sample. Consequently, the difference in the specific ULX frequency between the full and non-AGN sample could be explained by a non-linear dependence of the number

of ULXs on the $M_\star$.

In order to quantify the dependence of the specific ULX frequency on the $M_\star$, we compute the scaling factor b in our ETG sample, over three $M_\star$ bins ($10^{9.5} - 10^{10.0}$; $10^{10.0} - 10^{10.5}$; $10^{10.5} - 10^{11.0}$ $M_\odot$), separately for AGN and non-AGN galaxies. This is shown in Figure 3.5.4. The results for the AGN and non-AGN samples agree within the errors, as expected based on the previous assessment that the AGN do not lead to significant overestimation of the $M_\star$ (see Appendix 3.B). They also agree with the scalings reported in Zhang *et al.* (2012) and Plotkin *et al.* (2014) for the corresponding $M_\star$ bins (also plotted in Figure 3.5.4). Interestingly, however, we find that b depends strongly on the $M_\star$ of the host galaxy. This dependence explains the lower specific ULX frequency found in the 'full' ETG sample which is biased towards more massive galaxies than in the case of non-AGN ETG sample.

The dependence of the specific ULX frequency on the $M_\star$ could be caused by star-formation history (SFH) differences in ETGs (McDermid *et al.* 2015). Simulations indicate that ULXs with neutron-star accretors and red giant or Hertzsprung-gap donors can appear several hundreds of Myrs after a star-formation episode (Wiktorowicz *et al.* 2017), and therefore their frequency in early-type galaxies is expected to be strongly dependent on the SFHs. Calculated specific ULX frequencies (circles in Figure 3.5.4) using the Wiktorowicz *et al.* (2017) simulation and the McDermid *et al.* (2015) average SFHs for the same stellar mass ranges (see §3.5.5), are in excellent agreement with the observed specific ULX frequencies in our sample.

Therefore, comparisons of ULX rates in ETGs should account for the mass range covered in each sample and the corresponding bias due to different SFHs. In this respect, we attribute differences between our estimates of the specific ULX frequency and those of previous studies to the different $M_\star$ ranges in the samples. Note, however, that the specific ULX frequency was found to be constant in elliptical galaxies in Walton *et al.* (2011), albeit with a relatively small sample of 22 galaxies.

Furthermore, in §3.4.3 we find that lenticular galaxies in our sample host 2-3 times fewer ULXs than elliptical galaxies, by a factor of 2-3, even when normalising by the $M_\star$. This result is in agreement with the findings of Wang *et al.* (2016). However, we noticed that this difference disappears when considering the 'full' sample. Given the dependence of the specific ULX frequency on the $M_\star$ shown in Figure 3.5.4, it is possible that this discrepancy stems from the different $M_\star$ regimes of the corresponding samples. Indeed, we find that the interquartile range (middle 50%) of the stellar masses in the

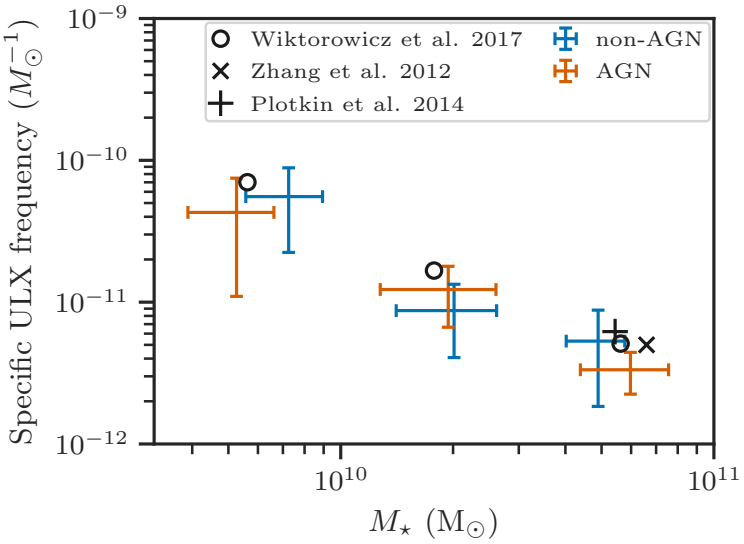


Figure 3.5.4: The specific ULX frequency for ETGs in our sample with stellar masses in the ranges $10^{9.5}$-10^{10}, 10^{10}-$10^{10.5}$ and $10^{10.5}$-10^{11} $M_\odot$, for AGN (orange error bars) and non-AGN galaxies (blue error bars). The cross and plus markers indicate the specific ULX frequency and the mean stellar mass of the samples in Zhang *et al.* (2012) and Plotkin *et al.* (2014). The circles indicate the computed values by convolving the binary population synthesis results of Wiktorowicz *et al.* (2017) with average SFH of ETGs in McDermid *et al.* (2015). Note that the x-axis error-bars do not indicate bin widths, but the standard deviation of the stellar masses of the galaxies contributing in each bin, to give a sense of the stellar mass distribution in each bin.

'full' sample lies in the $1.6{\times}10^{10}$ $M_\odot$-10^{11} $M_\odot$ range, for both elliptical and lenticular galaxies, while in the case of the non-AGN sample, lenticular galaxies present higher masses, $1.4\text{-}4{\times}10^{10}$ $M_\odot$, compared to that of the elliptical galaxies, $0.5\text{-}3.5{\times}10^{10}$ $M_\odot$.

3.5.5 Comparison with models

Comparison of binary population synthesis models and demographic studies of ULXs provide tests for models of the formation and evolution of X-ray binaries with extreme mass transfer rates.

We compare our findings with the results in Wiktorowicz *et al.* (2017), who computed the observed number of ULXs as a function of time for three different metallicities (0.01, 0.1 and 1 $Z_\odot$). Since the SFR indicators used in our study are based on the IR emission which is sensitive to stellar populations of ages up to ~100 Myr (Kennicutt & Evans 2012), we compare our results with the number of ULXs reported in

Wiktorowicz *et al.* (2017) observed after 100 Myr for a starburst scenario for 6×10^{10} $M_\odot$ of stars formed with 100 Myr duration. They report 4×10^2 ULXs which corresponds to a formation rate of 0.67 ULXs per $M_\odot$ yr^{-1}. This value is close to our results for all LTGs (0.51±0.06 ULXs per $M_\odot$ yr^{-1}; see Table 3.4.2).

To study the effect of metallicity, we also consider the 0.1 $Z_\odot$ simulation from Wiktorowicz *et al.* (2017). The resulting formation rate is 12 ULXs per $M_\odot$ yr^{-1}, about 18 times stronger than that in the case of Z=$Z_\odot$. As shown in Figure 3.5.3, our sample at such low metallicities is insufficient to estimate the excess of ULXs. However, we find an excess of ∼0.7 dex in the ULX rate for ∼0.3 dex lower metallicities in comparison to the bulk of the galaxies (which are predominantly solar metallicity galaxies). This translates to a factor of ∼5 more ULXs at Z=0.3$Z_\odot$ which is between the expectations from the models of Wiktorowicz *et al.* (2017) for Z=0.1 $Z_\odot$ and Z=$Z_\odot$.

However, the SFHs of real galaxies may present individual star-formation episodes (as can be the case in irregular galaxies). This is expected to have a strong effect on the formation rate of XRBs, as it has been demonstrated in HMXB populations (e.g., Antoniou & Zezas 2016; Antoniou *et al.* 2019; Lehmer *et al.* 2019), and the observed populations of ULXs which are typically associated to recent star-formation episodes. Although the SFHs of the galaxies in our sample are not known, based on the simulations of Wiktorowicz *et al.* (2017) we can estimate the range of ULX formation rates as a function of time in a continuous SF episode over a time-scale of 100 Myr. We compute the formation rate as the number of ULXs at time t from fig. 2 in Wiktorowicz *et al.* (2017) divided by the stellar mass of the parent population formed in the same SF episode. By performing this computation for various values of $t \leq 100$ Myr, we find formation rates in the range $0 - 1.58$ ULXs per $M_\odot$ yr^{-1} (see Figure 3.5.5) close to the range of the ULX-SFR scaling in samples of different morphological classes (see Table 3.4.2), except for the late-type spiral galaxies and irregulars (Sd-Im). The latter present an excess of ULXs due to their lower metallicities (see §3.5.2).

In the case of ETGs, most of the ULXs are expected to be long-lived systems of LMXBs with ages ≫100 Myr, but as shown in Wiktorowicz *et al.* (2017), the number of ULXs decreases with time since the SF episode. Therefore, the number of observed ULXs in ETGs, depends strongly on the SFH. In addition, recently, it has been shown in McDermid *et al.* (2015) that the age of stellar populations in ETGs can vary more than it was thought before. In order to compute a fiducial range of specific ULX frequency in ETGs, we use the average SFH of ETGs in McDermid *et al.* (2015) in three

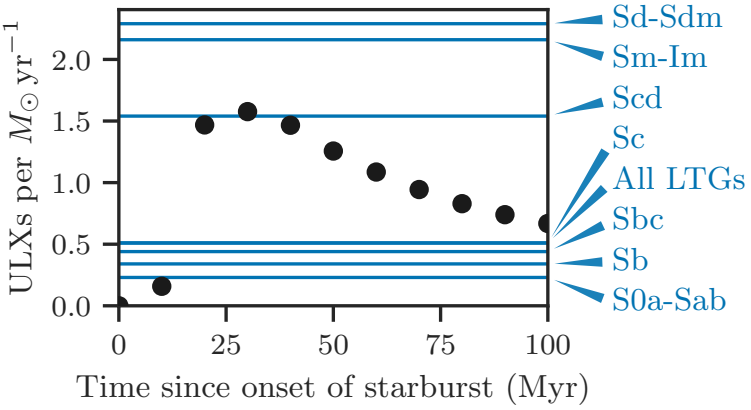


Figure 3.5.5: Formation rate of ULXs as a function of time since the onset of a star-formation episode (with constant SFR) based on the simulations of Wiktorowicz *et al.* (2017) for solar metallicity. The horizontal lines indicate the parameter a (number of ULXs per SFR) found in different morphological types of LTGs in our sample (see Table 3.4.2). The range of a in LTGs is comparable to the range of the computed formation rate of ULXs at timescales of $10 - 100$ Myr, except for Sd-Sdm and Sm-Im galaxies, possibly because of their sub-solar metallicity (see §§3.5.2, 3.5.3).

stellar-mass ranges: $\log \mathrm{M}_\star \in (9.5, 10.0)$, $(10.0, 10.5)$ and $(10.5, 11.0)$. These ranges cover the majority of the ETGs in our sample (∼90%).

For the three average SFHs, we compute the number of expected ULXs at the present time by convolving the SFHs (cf. Kouroumpatzakis *et al.* 2020) with the ULXs rates* per unit stellar mass from the prediction of Wiktorowicz *et al.* (2017) for solar metallicity. Then, we divide by the the midpoint for each mass range in order to calculate the specific ULX frequencies in each mass range. We find 70.1, 16.7 and 5.1 ULXs per $10^{12}\mathrm{M}_\odot$ for the low-, intermediate- and high-mass ETGs, suggesting that the ULX content of ETGs is indeed a strong function of SFH (see Figure 3.5.4). These estimates are comparable to the rates we derived from our ETGs sample, b=$15.1^{+3.9}_{-3.6}$ ULXs per $10^{12}\mathrm{M}_\odot$ for non-AGN ETGs, and b=$6.3^{+1.0}_{-0.9}$ ULXs per $10^{12}\mathrm{M}_\odot$ for the full ETGs sample (which is biased towards the higher mass bin; see Table 3.4.2 and Appendix 3.B).

* Since Wiktorowicz *et al.* (2017) do not provide an instantaneous SB response function, rather a SB of duration of 100 Myr, the convolution is performed in bins of 100 Myr.

3.5.6 Limitations of this study

The parent sample of the *HECATE*, the *HyperLEDA*, includes galaxies and measurements from a multitude of surveys with different sky coverage and sensitivity. Similarly, parameters provided in *HECATE* (e.g., SFR, $M_\star$, metallicity, AGN classifications) are derived from combinations of data from all-sky surveys (e.g. *IRAS*, *2MASS*) and the *SDSS* (e.g., M/L ratios, *WISE* forced photometry of *SDSS* objects). Despite the unknown selection function of the parent sample, the *HECATE* is the most complete sample of galaxies in the local Universe with available information on their stellar content, allowing us to draw meaningful conclusions regarding the ULX scaling relations covering a very broad stellar mass (10^7.5-$10^{11.5}$ $M_\odot$) and SFR range ($10^{-2.5}$-10^2 $M_\odot$ yr^{-1}; §3.3.3.2).

Similarly, the serendipitous nature of the *CSC 2.0* leads by definition to a non-uniform X-ray sample. In addition, to avoid contamination from X-ray emitting AGN, we exclude nuclear sources in galaxies that were either classified as AGN or we did not have information on their nuclear activity. The drawback of this approach is that we may have excluded circum-nuclear ULXs.

Finally, for the study of scaling relations, we primarily consider a secure sample of non-active galaxies to avoid the overestimation of SFR and $M_\star$ due to nuclear activity. This practice reduces the sample used for the ULX investigations, and may have removed known bona-fide starforming ULX hosts (e.g., Holmberg II). In addition, it leads to a bias against massive galaxies which are more likely to be targeted as AGN hosts. As discussed in §3.5.4, including the AGN sample, at least in the ETGs, does not bias the measured galaxy properties.

3.6 Summary

We construct a census of ULXs in nearby galaxies (D<40 Mpc) by cross-matching the *CSC 2.0* and the *HECATE*. We use this sample in order to study the ULX rates as a function of morphology, SFR, $M_\star$ and metallicity of their host galaxies. We deliver a sample of host galaxies and their ULX populations that serves as a benchmark for models describing the nature, formation and evolution of ULXs. We

(i) constrain the number of ULXs in LTGs as a function of SFR, and both SFR and $M_\star$ (to account for the LMXB contribution):

$$N_{\rm ulx} = (0.51 \pm 0.06) \times \frac{\rm SFR}{\rm M_\odot\,yr^{-1}} \tag{3.11}$$

$$N_{\rm ulx} = 0.45^{+0.06}_{-0.09} \times \frac{\rm SFR}{\rm M_\odot\,yr^{-1}} + 3.3^{+3.8}_{-3.2} \times \frac{M_\star}{10^{12}\,\rm M_\odot}. \tag{3.12}$$

(ii) find that the ULX-SFR scaling increases with the morphological type of LTGs.

(iii) verify the excess of ULXs in low-metallicity galaxies, which partially drives the above mentioned trends with the morphological type.

(iv) find evidence for evolution of the specific ULX frequency in ETGs with their $M_\star$, which we attribute to their different SFHs.

(v) find that our observed scaling relations can be reproduced by published ULX formation rates from population synthesis models when accounting for the galaxies SFHs and/or metallicity.

While *eROSITA* (Predehl *et al.* 2010; Merloni *et al.* 2012) will provide a uniform flux-limited sample of normal galaxies and ULXs in the local Universe (e.g., Basu-Zych *et al.* 2020), serendipitous surveys with *Chandra* will continue to probe unconfused ULX populations at larger distances and their connection to the lower luminosity XRB populations.

Appendix

Appendix 3.A Galactocentric scale parameter

We define the galactocentric scale parameter, c, as the deprojected distance of a source from the centre of its host galaxy, normalised by the galaxy's semi-major axis. Since the shapes of the galaxies in our study are defined through isophotal ellipses and the length of the semi-major axis is free of projection effects, we observe that c can be computed as the ratio of the semi-major axes of two projected ellipses: a scaled version of the isophotal ellipse (same centre, orientation and axis ratio) passing through the source, and the isophotal ellipse itself. Consequently, for a source at (α_s, δ_s) and a galaxy centred at (α_g, δ_g) with semi-axes R_1, R_2 and position angle ω measured from North to East, the scale c is found by

(i) rotating the coordinates so that the centre of the galaxy falls in (α, δ)= $(0, 0)$ and the semi-major axis is on a meridian (a=0)

(ii) setting the sum of the great-circle distances of the source from the focal points to be equal to two times the semi-major axis (of the scaled version of the ellipse)

Step (i) is performed by converting the spherical coordinates to Cartesian (unit radius):

$$(x, y, z) = (\cos\delta_s \cos\alpha_s, \cos\delta_s \sin\alpha_s, \sin\delta_s)\,, \tag{3.13}$$

and rotating around the z-axis by $-\alpha_g$, the y-axis by δ_g and the x-axis by $-\omega$ (to align the semi-major axis with the meridian), by multiplying with the corresponding

3D-rotation matrices:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{R}_x(-\omega) \cdot \mathbf{R}_y(\delta_g) \cdot \mathbf{R}_z(-\alpha_g) \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{3.14}$$

The final coordinates are converted back to spherical coordinates:

$$(\alpha, \delta) = \left(\tan^{-1} \frac{y'}{x'}, \frac{\pi}{2} - \cos^{-1} z' \right). \tag{3.15}$$

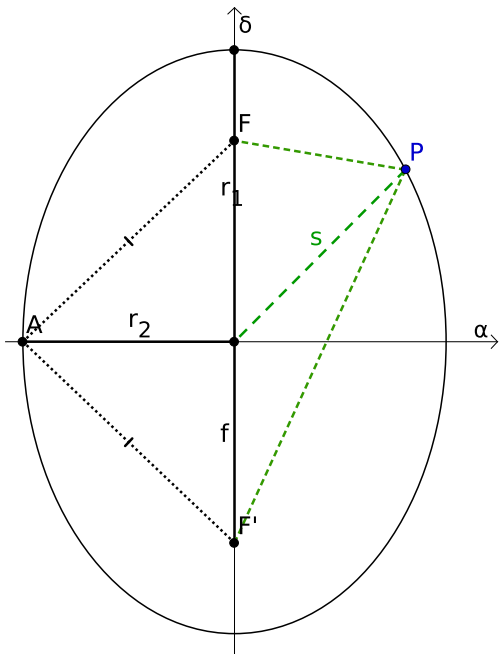


Figure 3.A.1: The scaled version of the ellipse passing through the source (P). The separation of the source s and its distance from the two focal points (F, F') is denoted with green dashed lines. The solid black lines denote the semi-major and semi-minor axes, while the dashed black lines indicate the distance of the co-vertex (A) to the focal points. The co-vertex A is introduced so that the focal distance f is estimated in an intermediate step.

Step (ii) consists of finding the parameters of an ellipse shown in Figure 3.A.1 for which the semi-major and semi-minor axes are scaled versions of the original ellipse $(r_1, r_2) = (cR_1, cR_2)$. This is done by requiring $(F, P) + (F', P) = (F, A) + (F', A)$ where (A, B) denotes the great-circle distance between points $A\,(\alpha_1, \delta_1)$ and $B\,(\alpha_2, \delta_2)$, computed by employing a form of the Haversine formula which is more precise for nearby points:

$$s = 2 \sin^{-1} \sqrt{\sin^2 \frac{\delta_1 - \delta_2}{2} + \cos \delta_1 \cos \delta_2 \sin^2 \frac{\alpha_1 - \alpha_2}{2}}. \tag{3.16}$$

The solution in terms of c (the scale of the ellipse) is found by solving for c the equation:

$$\sin^{-1}\sqrt{u^{-}+w}+\sin^{-1}\sqrt{u^{+}+w}=cR_1, \tag{3.17}$$

where

$$u^{\pm}=\sin^2\frac{f\pm\delta}{2}, w=\cos f\cos\delta\sin^2\frac{\alpha}{2}, f=\cos^{-1}\left[\frac{\cos\left(cR_1\right)}{\cos\left(cR_2\right)}\right]. \tag{3.18}$$

Since the Equation 3.17 is not in closed form, it is solved numerically. Due to the periodicity of trigonometric functions, there are multiple solutions corresponding to ellipses engulfing the celestial sphere multiple times. To avoid this, we require that r_2 is less than $\frac{\pi}{2}$. Also, the separation of the source from the centre of the galaxy acts as a lower and upper limit for the semi-major and semi-minor axes respectively. Therefore the galactocentric distance is constrained on

$$c\in\left[\frac{s}{R_1},\min\left\{\frac{s}{R_2},\frac{\pi}{2R_1}\right\}\right], \tag{3.19}$$

where s is now expressed in the transformed coordinates as:

$$s=2\sin^{-1}\sqrt{\sin^2\frac{\delta}{2}+\cos\delta\sin^2\frac{\alpha}{2}}. \tag{3.20}$$

Appendix 3.B Results of fits with and without AGN hosts

As described in §3.3.3.2, the far- and near-infrared emission from AGN may bias the estimates of SFR and $M_\star$ in our sample. Therefore, when fitting the models in §3.4.3, i.e. number of ULXs as a function of SFR and $M_\star$ in late- and early-type galaxies, we considered only galaxies that were classified as non-AGN in the *HECATE*. Here, we perform the same analysis for the complete sample (including the AGN and unclassified galaxies). The results of these fits also enable the direct comparison of this study with previous works where the AGN-hosts were not excluded from their samples.

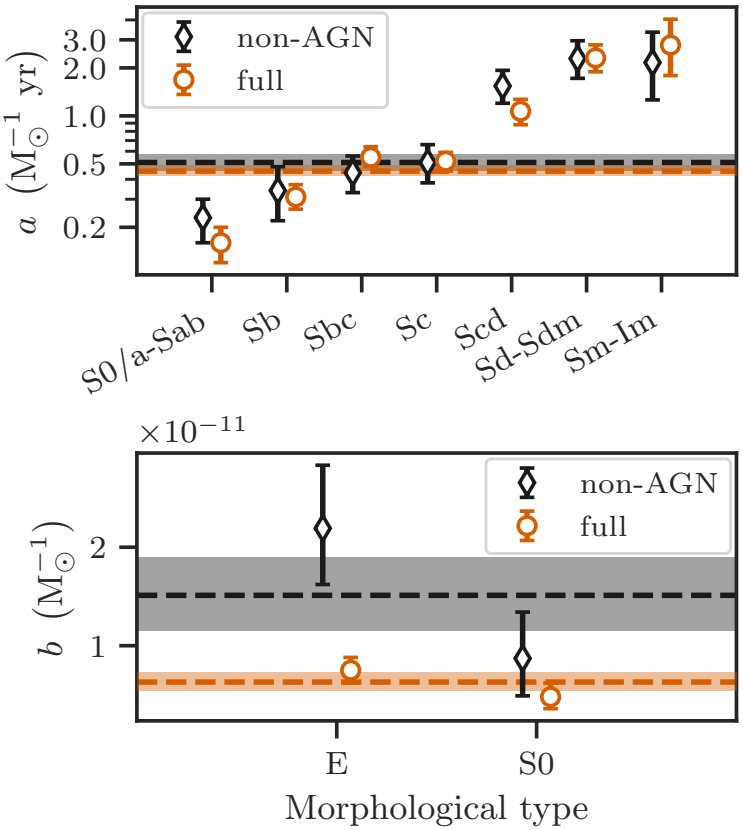


Figure 3.B.1: Comparison of fitting results between non-AGN and 'full' samples. **Top**: the scaling parameter a (see Equation 3.8) in non-AGN LTGs (dashed black line) and its 68% CI (grey band), and in the 'full' sample (orange line and band). Fitting results for various morphological subclasses of LTGs are shown with error bars. **Bottom**: same as top panel, but now for the scaling parameter b (see Equation 3.7) in ETGs and separately in elliptical and lenticular galaxies.

In the following paragraphs, the complete sample without removing any AGN hosts is referred as 'full' sample. The sample used in §3.4 and §3.5, where galaxies with nuclear activity or lacking classification in the *HECATE* were excluded from the fits, is referred as 'non-AGN'. The 'full' sample is larger than the 'non-AGN' sample by a factor of ~4 in elliptical, ~3 in lenticular, ~2 in spiral and ~ 1.2 in irregular galaxies.

The top panel of Figure 3.B.1 shows the scaling of the number of ULXs with the SFR (parameter a in Equation 3.8) in LTGs (top left) and their morphological sub-classes (top right), for the full (black) and the 'non-AGN' (orange) samples. We find $a = 0.45^{+0.04}_{-0.03}$ ULXs per $M_\odot$ yr^{-1} in the case of the 'full' sample, lower than that

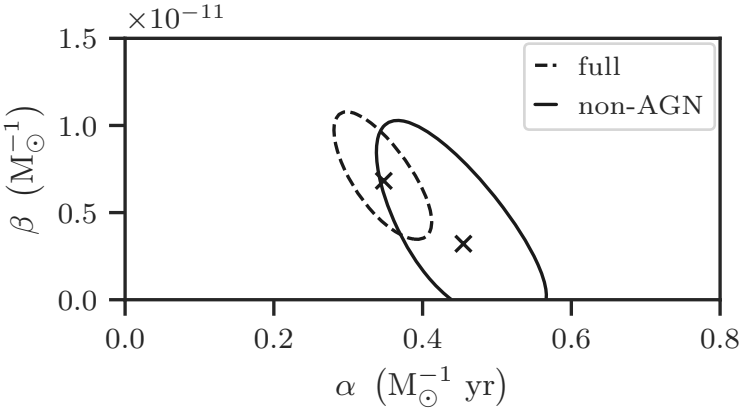


Figure 3.B.2: The best-fitting values (X symbols) and, 68% CIs (lines) of the scaling parameters α and β of Equation 3.9, for all late-type galaxies (solid) and their non-AGN subset (dashed). The bias due to the overestimation of SFR in AGN hosts, manifests as an underestimation of the scaling parameter α. The value of β is highly uncertain to notice any bias due to the inclusion of AGN (lower contribution of old stellar populations in the number of ULXs in LTGs).

of the 'non-AGN' group, 0.51 ± 0.06 ULXs per $M_\odot$ yr^{-1}. However the difference is not significant ($\sim1\sigma$). The comparison between the 'non-AGN' and 'full' sample for the ULX-SFR scaling in different morphological types is not conclusive because of the large uncertainties (see Figure 3.B.1).

The bottom panel of Figure 3.B.1 shows the posterior probability distribution of the scaling of the number of ULXs with $M_\star$ (b) for the 'full' (solid line) and 'non-AGN' (dashed line) ETGs. We find that the specific ULX frequency is significantly lower when 'full' ETGs are considered ($6.3^{+1.0}_{-0.9}$ ULXs per $10^{12}M_\odot$) than in the case of 'non-AGN' ETGs ($15^{+3.9}_{-3.6}$ ULXs per $10^{12}M_\odot$), with similar results between elliptical and lenticular galaxies.

For the scaling of the number of ULXs with SFR and $M_\star$ in LTGs, the posterior distribution of the two scaling factors (α and β) considering the 'full' sample is shown in Figure 3.B.2. For comparison we also show the 'non-AGN' case (dashed lines; presented in §3.4.3). We find for the 233 galaxies in the 'full' sample, hosting 328 ULX candidates (48.7 expected f/b sources), $\alpha = 0.35^{+0.04}_{-0.05}\,M_\odot\,\mathrm{yr}^{-1}$ and $\beta = 6.8^{+2.6}_{-2.2}\times10^{-12}$ ULXs per $10^{12}M_\odot$ (marginalised). The difference of the posteriors for the scaling factor α is consistent with the difference seen in the fits with SFR scaling only.

Given their broad-band SED, AGN may have significant contribution in the optical/UV of their host galaxies, and also in their FIR emission in the case of type-2 AGN (cf. Risaliti & Elvis 2004). However, their contribution in the near-infrared part of the spectrum is relatively small. Therefore, the difference we find in the specific ULX frequency between the 'full' and 'non-AGN' samples of ETGs is unlikely to be due to an overestimation of the K-band based $M_\star$ estimates in possible AGN in the full sample.

This is supported by the X-ray luminosities of the nuclear sources in the full ETG sample. These are very low (only one exceeds 10^{41} erg s^{-1}), indicating that if there is an AGN its contribution to the K-band luminosity may not be significant. Additional support comes from the ratio of the nuclear and the total K-band luminosity in a representative sample of star-forming galaxies of Bonfini et al. (2020; submitted). This study (based on morphological decomposition of K-band images of the Star-Formation Reference Survey; Ashby *et al.* 2011) shows that the typical AGN contribution to the galactic K-band luminosity is <10%. We conclude that the $M_\star$ estimates in AGN-hosts are not expected to be significantly biased upwards by the potential presence of an AGN.

Therefore, as discussed in §3.5.4, the difference in the specific ULX frequency in the non-AGN and the 'full' ETG samples is the result of the SFH differences in ETGs of different stellar masses.

Appendix 3.C Detailed comparisons with previous ULX surveys

In the following subsections we compare our results with two major surveys of ULXs in the local Universe: Swartz *et al.* (2011) (hereafter S11) and Wang *et al.* (2016) (hereafter W16).

3.C.1 Comparison with *Swartz et al. (2011)*

Based on the total late-type galaxy sample in our study, we find a SFR scaling factor of ~0.51±0.06 ULXs per $M_\odot$ yr^{-1} (see Table 3.4.2) which is ~4 times lower than that found by S11 (~2 ULXs per $M_\odot$ yr^{-1}).

The main reason behind this discrepancy is the different samples of host galaxies in terms of morphology: S11 consider a large population of late-type and irregular galaxies, as shown in their fig. 1). Using the reported fractions of morphological types in their sample, and our fitting results for a in different sub-classes of LTGs (see Table 3.4.2), we compute the expected number of ULXs in a sample with the same distribution of morphological types as in S11, based on our findings. We find 1.2 ULXs per $M_\odot\ yr^{-1}$, a factor of 2.4 higher than in our sample. Another reason for this discrepancy could be differences in the X-ray photometry of the sources*. Indeed, for the 99 common X-ray sources (cross-match radius of 3″), only 73 of them are characterised as ULX candidates by us, while we find that the luminosities we report are smaller by ~ -0.4 dex (a factor of ∼0.4) compared to those computed by S11. Assuming a cumulative slope of 0.6 (appropriate for the HMXB-LF; Grimm *et al.* 2003; Mineo *et al.* 2012), this corresponds to a lower number of ULXs by a factor of ∼1.7 compared to S11. In total, the combination of the two factors give a factor of 4.2 lower estimate of the SFR scaling factor in our sample, explaining the difference we find.

Finally, in our sample, we find $\sim 0.68 \pm 0.10$ ULXs per elliptical galaxy (see Table 3.4.1) which is significantly higher than the rate (0.23) reported in S11. We attribute this discrepancy to the small number statistics, and the under-representation of elliptical galaxies in S11. On the other hand, the fraction of elliptical galaxies in our sample is similar to that of W16, which presents comparable number of ULXs per elliptical galaxy (0.43±0.11).

3.C.2 Comparison with *Wang et al. (2016)*

We cross-check the number of ULXs in our sample against the results of W16, the largest and most recent study of ULX demographics with *Chandra* observations. To do so, we compare the total number of ULXs in the common galaxies in our sample and in that of W16.

As a first step, we cross-match the two galaxy samples. Out of the 343 galaxies in the sample of W16, 315 are associated with our host galaxy sample. The remaining 28 galaxies are not included in our sample for various reasons. In 22 cases, the targets

* S11 used a count-rate to flux conversion factor assuming an absorbed power-law model with $\Gamma = 1.8$ and, for sources with > 130 counts, performed spectral fits or adopted published results. Instead, we adopt the aperture-corrected net energy flux from the *CSC 2.0*.

were observed with shallow observations (exposure times $\lesssim 5$ ks) and *Chandra* did not detect any source. In the remaining six cases, the sources in W16 do not lie in the D_{25} regions of the galaxies, our criterion for associating sources to host galaxies (NGC3066, NGC1507), or the observations were not included in the *CSC 2.0* (NGC3489, NGC3489, PGC48179, PGC35286).

However, there are three additional important differences between this study and that of W16 that need to be considered in the comparison:

(i) In W16, X-ray sources must exceed $2{\times}10^{39}$ erg s^{-1} in X-ray luminosity to be considered as ULX candidates, instead of our criterion of $L_{\rm X}{>}10^{39}$ erg s^{-1}.

(ii) The X-ray source sample of W16 was taken from Liu (2011) who computed the fluxes of the sources using a count-to-flux conversion assuming a power-law spectrum with Γ=1.7 and Galactic line-of-sight absorption. Instead, we use the net energy of the photons as reported in the *CSC 2.0* (see §3.3.5).

(iii) In this study, we associate sources to a host galaxy if they lie in its D_{25} region. In W16 the $2{\times}D_{25}$ regions are used, namely the sky ellipses with twice the major and minor axes.

In order to account for the luminosity difference, we cross-match our X-ray sample with the one used by W16, and we find that our luminosities are 6% smaller (median ratio; scatter of 0.5 dex). Therefore, for this comparison only, we will consider as ULXs in our sample, the 'reliable' X-ray sources with $L_{\rm X} > 1.89 \times 10^{39}$ erg s^{-1} to account for the above mentioned luminosity offset. We find a total of 186 ULX candidates in our sample. The same luminosity limit is used to calculate the expected foreground/background contamination (see §3.3.6), for which we find 37.9 sources, leading to a final estimate of 148.1 ULXs in the 315 galaxies, based on our emulation of the W16 sample.

To correct for the difference between the number of sources in the D_{25} and $2{\times}D_{25}$ regions, we cross-match the X-ray sample with the *HECATE* and find that the number of X-ray sources that lie in the $2{\times}D_{25}$ regions is 35% larger than the number of sources in the D_{25} regions. Therefore, the number of ULX candidates in the $2{\times}D_{25}$ regions, 215, as reported in W16, corresponds to about 159.3 candidates in the D_{25} regions. By subtracting one forth of the, reported by W16, number of interlopers (since the area of the D_{25} regions is four times than the area of the $2{\times}D_{25}$ regions), we find that the

number of ULXs in the D_{25} regions is 151.1, close to the value we find in our sample (148.1).

Lastly, the ULX rates above 2, 3, 5 and 10×10^{39} erg s^{-1} in the different morphological types in our sample (see Table 3.4.1) are consistent at the 2σ-level with the rates found in the table 2 in W16.

This page intentionally left blank

4

Source confusion and ULXs

4.1 Background

Source confusion (or blending) is an inherent problem in astronomical surveys. Its magnitude depends on the telescope resolution, and the spatial distribution of the sources. Due to the relatively low resolution of X-ray observatories, source confusion limits the distance at which sources can be probed in external galaxies. While luminous sources like active galactic nuclei (AGN) may dominate the output of their host galaxies and can be studied even at high redshift, sources of lower luminosities, such as X-ray binaries (XRBs), may be blended in nearby galaxies.

In the case of low-mass XRBs, which follow the mass distribution of galaxies, confusion is expected to be more significant in the bulges and circum-nuclear regions of galaxies. On the other hand, the confusion of high-mass XRBs, which are associated to star-forming regions, is dictated by the clumpy and hierarchical nature of star-formation (e.g., Elmegreen & Elmegreen 2001). Consequently, the magnitude of the confusion effect may vary depending on the morphological type of the galaxies, and hence their 'clumpiness' (e.g., Elmegreen *et al.* 2009).

If the effect is significant, the numerous sources in the lower-end of the luminosity function (LF), will have a higher chance of being blended into other sources of comparable or higher luminosities. Therefore, the numbers of faint sources will decrease, and additional brighter sources will appear in the luminosity function. The overall effect will be seen as a flattening of the LF, and a decrease in the total number of sources (i.e., normalisation constant).

Source confusion limits the study of XRBs as individual objects to the Milky Way and nearby galaxies ($D \lesssim 10$ Mpc). This introduces major challenges in demographic surveys of XRBs: the lower the distance, the smaller the sample size, and the variety of the hosting galactic environments. For example, the Galactic neighbourhood is dominated by late-type galaxies, with massive elliptical galaxies being encountered usually at more distant galaxy clusters (>16 Mpc). Consequently, studies of X-ray binary populations in the context of the stellar population parameters of their host galaxies, and the respective scaling relations (e.g., Grimm *et al.* 2003; Gilfanov 2004; Mineo *et al.* 2012; Zhang *et al.* 2012; Brorby *et al.* 2014; Lehmer *et al.* 2019) are limited to a narrow parameter space (in star-formation rate, stellar mass, and metallicity).

The above mentioned surveys are important for: (i) understanding the formation and evolution of XRBs; (ii) understanding the connection with the star-formation history and metallicity of the host galaxies; and (iii) constraining stellar and binary evolution theory. The quantification of the source confusion effect on the LFs, demographics and scaling relations, is crucial for fully exploiting data from previous and current observations (e.g., *Chandra*, *XMM-Newton*), as well as upcoming surveys (e.g., *eROSITA*, *Athena*).

4.1.1 Addressing the source confusion problem

The obvious way to reduce the source blending effect in a survey, is to use a high-resolution telescope for the observations. For this reason, studies of XRBs in external galaxies are usually performed using the *Chandra* X-ray observatory (e.g., Fabbiano *et al.* 2006).

In demographic studies, an optimal distance limit is adopted for minimising the effect of confusion, while covering an adequate sample of galaxies. For example, studies focusing on bright HMXBs (e.g., $L_X \gtrsim 10^{38}$ erg s^{-1}), and ultraluminous X-ray

sources (ULX; by definition $L_X > 10^{39}$, erg s^{-1}), usually choose a distance limit of 40 Mpc for the galaxy sample (e.g. Mineo *et al.* 2012). This limit was also our choice in §3.

One way to quantify the effect of confusion in the LF of XRBs (especially in studies of even more distant galaxies: $D \gtrsim 70$ Mpc) is to perform a simulation of X-ray data constructed by distributing artificial XRBs (from canonical LFs) according to the star-formation rate map (or a proxy of it) for a given galaxy. Following this procedure, Basu-Zych *et al.* (2016) tested whether the observed excess of ULXs in local ($D \approx 85$ Mpc) Lyman-break analogs can be attributed to confusion effects or their low-metalliciy which is known to lead in increased numbers of ULXs (e.g. Mapelli *et al.* 2010). This approach is not practical for large galaxy samples, since it relies on modelling each galaxy separately.

In this work, we aim at quantifying in a statistical way, and as a function of distance, the effect of confusion in the luminosity functions of XRBs, and especially the population of ULXs, which have been the focus of our recent study using *Chandra* data (see §3). In §4.2 we describe the methodology for simulating X-ray observations of galaxies at different distances, the analysis of the simulated data, and the confusion metrics we use. In §4.3 we present the results, and in §4.4 we discuss the implications of these findings for demographic studies of ULXs.

4.2 Methodology

To study the effect of source confusion in X-ray observations, we obtain X-ray images of nearby galaxies ($D < 10$ Mpc). Then, by binning them at different scales, we will simulate the conditions of observing them at larger distances. By detecting the X-ray sources in the *original* and *simulated* distances, we will obtain the necessary data for exploring the effect of source confusion in the luminosity distributions of X-ray sources. This data will be used to estimate the distance at which ULX populations remain virtually unaffected by source blending.

4.2.1 The galaxy sample

In order to investigate the source confusion as a function of the distance, we require a sample of nearby galaxies for which we can obtain a relatively complete

Table 4.2.1: Basic parameters of the galaxy sample.

Name	RA	Dec	Type	R_1	R_2	D	Coverage	$f_{\rm min}$	N
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
M51	202.46955	47.19515	SABb	6.9	5.8	8.6	1.00	2.8	243
M63	198.95540	42.02928	Sbc	5.9	4.0	8.9	1.00	29.7	62
M83	204.25390	-29.86563	Sc	6.8	6.5	4.8	1.00	2.7	319
M101	210.80237	54.34902	SABc	12.0	11.5	6.9	1.00	1.1	246
NGC 55	3.72337	-39.19658	SBm	14.9	1.5	1.8	0.91	15.1	68
NGC 300	13.72302	-37.68448	Scd	9.7	6.5	2.0	0.99	7.2	131
NGC 2403	114.21424	65.60242	SABc	10.0	5.0	3.3	0.98	6.2	91
NGC 2903	143.04205	21.50157	Sbc	6.0	2.6	9.4	1.00	15.1	58
NGC 4449	187.04626	44.09363	IB	2.3	1.4	4.7	1.00	20.0	26
NGC 5253	204.98330	-31.64004	SBm	2.5	1.1	3.8	1.00	9.7	21
NGC 6946	308.71917	60.15352	SABc	5.7	5.4	5.5	1.00	4.0	129
NGC 7793	359.45759	-32.59107	Scd	5.2	3.0	3.4	1.00	7.3	36

Column descriptions: (1) Galaxy in Messier or NGC catalogue; (2) Decimal J2000.0 right ascension in degrees; (3) Decimal J2000.0 declination in degrees; (4) Morphological type; (5) D_{25} semi-major axis in arcmin; (6) D_{25} semi-minor axis in arcmin; (7) Adopted distance in Mpc; (8) Fraction of D_{25} area covered by *Chandra* observations; (9) Flux of the faintest source in 10^{-16} erg s^{-1} cm^{-2} at the D_{25} region in the original image; (10) Number of X-ray sources at the D_{25} region in th original image.

census of their XRB populations. We select galaxies from the host galaxy sample presented in §3, requiring they meet the following criteria:

1. distance less than 10 Mpc, to avoid source confusion effects at the original distance;
2. larger than 90% coverage in the D_{25} region, to avoid over-representation of specific regions (e.g., bulges, disks);
3. at least 20 X-ray sources detected at the original distance, to avoid small number statistics;
4. morphological type later than Sb, to study the actively star-forming galaxies;
5. galactic latitude $|b|>$ 10 deg, to avoid significant galactic absorption;

In addition, we exclude edge-on galaxies with strong galactic winds, since diffuse emission clumps will significantly contaminate the population of detected X-ray sources. In total, we select 12 galaxies, the basic parameters of which are listed in Table 4.2.1.

4.2.2 The X-ray source sample

The X-ray data needed for this study, are the luminosity distributions of X-ray sources in each galaxy at the original distances, as well as simulated distances. We adopt the following set of distances: 5, 10, 20, 30, 40, 50, 60, 80, and 100 Mpc. This range of distances covers the full range of our ULX census in §3 and extends more than twice the distance limit of 40 Mpc which is commonly adopted in HMXB and ULX demographics.

Firstly, we obtain X-ray images from the *Chandra* archive for each galaxy. The analsis of the data is performed with `ciao 4.12`. First, the event files are spatially binned at different scales to create images emulating the observation of galaxies at different distances. We use `wavdetect` to detect sources in these images. The source detection is performed on scales of 2, 4, and 8 pixels in order to minimise contamination by extended diffuse emission clumps, while ensuring the detection of sources broadened by the effect of PSF at the edges of the field for large, extended galaxies. In the case of galaxies with 5 Mpc$<D<$10 Mpc, we do not produce images at the simulated distance of 5 Mpc. By performing source detection on the binned images we produce source lists for each galaxy, at each simulated distance. In our analysis we only use sources with signal-to-noise ratio above 3. Figure 4.2.1 shows the galaxy M83, along with the positions of the sources detected at the minimum (5 Mpc) and maximum distance (100 Mpc).

In order to convert the count rate to luminosity, we assume a power-law spectrum with photon index $\Gamma = 1.7$, and line-of-sight column density $N_{\rm H}$=10^{20} cm^{-2}. While the assumed model will affect the absolute luminosity of the detected sources, it will not affect the comparison between the simulated sources for the different distances. In Appendix 4.A, we present the resulting luminosity distributions of the sources for each galaxy and distance.

Demographic studies of ULXs often exclude nuclear regions from their analysis (e.g., Swartz *et al.* 2011; Wang *et al.* 2016; Kaaret *et al.* 2017; Kovlakas *et al.* 2020). In order to provide insights for similar studies, we also construct LFs excluding nuclear sources, i.e., with separation less than 3 arcsec from the centre of the galaxy in each simulation, following our criterion in §3.3.7. The LF of *off-nuclear* sources are shown in Figure 4.A.2.

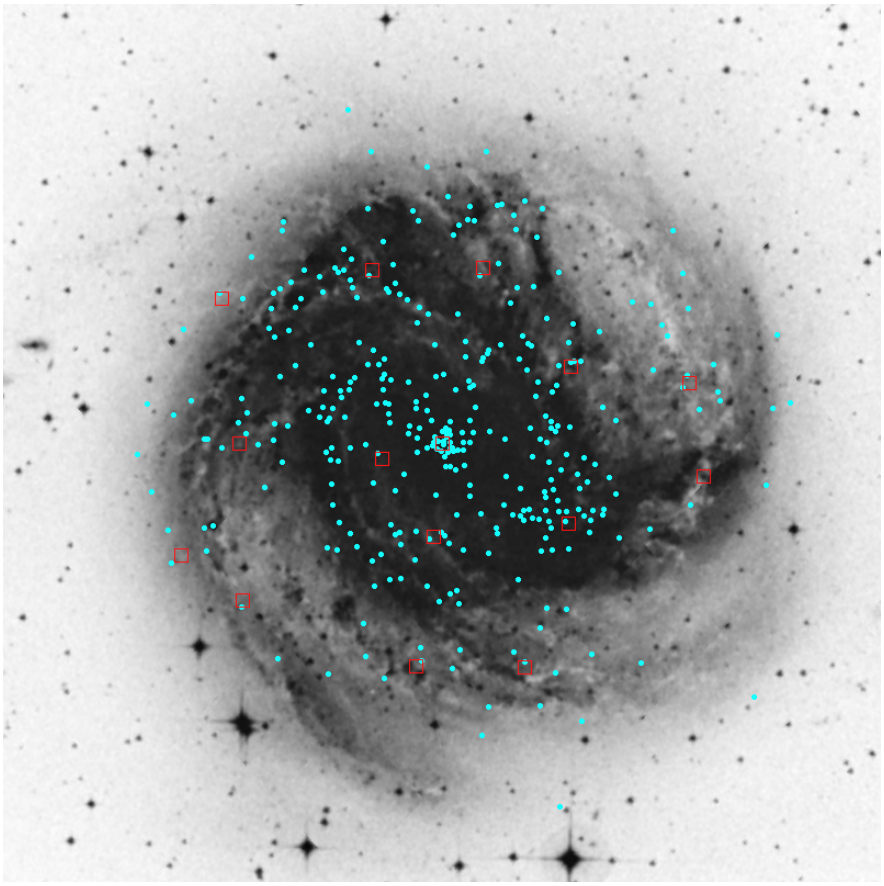


Figure 4.2.1: DSS-blue image of M83 and the X-ray sources in its D_{25} region, detected for the original distance of 4.8 Mpc (cyan filled circles) and the simulated distance of 100 Mpc (red squares). The number of sources at 100 Mpc is relatively small, and are characterised by the joint emission of multiple sources (at the original distance), especially at regions of high source density (e.g., the nucleus).

4.2.3 Source confusion metrics

To quantify the effect of source confusion we use three metrics, presented in the following paragraphs.

4.2.3.1 Flattening of the luminosity functions

We probe the expected flattening of the LFs by studying the evolution of the slope of the LF with distance. We assume a simple power-law model, $\frac{dN}{dL} \propto L^{-a-1}$. The cumulative slope, a, and its uncertainty, are estimated though the maximum-likelihood formulas of Crawford *et al.* (1970).

The calculated slopes and uncertainties are used to test for monotonicity in different ranges of simulated distances. We perform ordinary least square fits of the distance vs. slope data, where the slopes are Monte Carlo sampled from the Gaussian distributions with means and standard deviations equal to the estimated slopes and their uncertainties. If the 90% of the linear slopes are found negative (positive) we consider that the LF slope at the specific distance range is decreasing (increasing).

4.2.3.2 Number of ULXs as a function of distance

Since we are interested in the ULX regime of the LFs, we cannot rely on the computation of slopes above 10^{39} erg s^{-1} due to the small number of luminous sources. For this reason, we directly count the number of ULXs in each sample galaxy. Essentially, a change in the number of ULXs with distance in each galaxy, indicates the distance at which source blending affects their number.

4.2.3.3 Confusion distance as a function of luminosity

While the first metric (power-law slope) accounts for all sources in the LF, it may be biased by deviations from the simple power-law, but more importantly it does not probe the confusion effect at different luminosities. On the other hand, the second metric (number of ULXs) while focusing on the ULX regime, it may be affected by small number statistics. For this reason, we define a new metric, the *confusion distance* which combines advantages of both of the above metrics.

Specifically, *the confusion distance at luminosity L in a given galaxy, is the distance d_C so that at any distance above d_C, the number of sources with luminosities above L differs from the number of sources above the same threshold at the original distance, by at least two.* The value of two is selected so that the metric is not sensitive to stochastic effects that may produce small fluctuations. Finally, assuming different luminosity thresholds, and combining the estimates of d_C for all galaxies in our sample, we estimate the average confusion distance as a function of the source luminosity.

4.3 Results

4.3.1 Slopes of luminosity functions

Figure 4.A.3 shows the slopes of the LFs of all sources in each galaxy, at each simulated distance. Based on the results of the monotonicity tests (see §4.2.3.1), we find a trend for flattening of the LF slope as a function of distance for three galaxies (M101, M83, and NGC 2403). Fore another four galaxies (NGC 6946, NGC 7793, NGC 55, and NGC 300) we do see a similar trend, which however is not systematic: it appears mostly when comparing slopes at D<40 Mpc with slopes at D>40 Mpc. In four galaxies (NGC 4449, NGC 5253, M51, and M63) we do not see any evidence for significant change in the slope within the uncertainties. Interestingly, we find an increase of the slope with distance, in NGC 2903 between D = 10 Mpc and D>50 Mpc.

The above findings do not change when excluding the nuclear sources (see Figure 4.A.4), as it is expected by their small contribution in the overall populations, and consequently their negligible effect in the computation of the slopes.

4.3.2 Number of ULXs as a function of distance

We count the sources with luminosities above 10^{39} erg s^{-1} (henceforth, $N_{\rm ULX}$) in each galaxy in our sample, and for all distances (see left panel of Figure 4.3.1). In half of the galaxies (M83, NGC 2403, NGC 4449, NGC 5253, NGC 6946, and NGC 7793) there is an increase of $N_{\rm ULX}$ at various distances in the range of 10–100 Mpc. On the other hand, in the ULX-rich galaxies, M51, M63, M83, and NGC 2903, we notice fluctuations in the number of ULXs as a function of distance: they initially decrease at 10–40 Mpc, but increase at 80–100 Mpc. In the remaining galaxies, NGC 55 and NGC 300, there is no ULX found even at large distances.

The above trends are also observed when excluding the nuclear regions of the galaxies (see right panel of Figure 4.3.1), with the following differences: (i) the fluctuations in $N_{\rm ULX}$ are reduced, (ii) the extra ULXs appearing in M83 (at D=100 Mpc), NGC 2903 (at D=100 Mpc), and NGC 4449 (at D=60 Mpc) are not found anymore, (iii) no ULX is found in NGC 5253, NGC 6946.

The mean distance at which the number of ULXs increases, and its standard error, is 63±9 Mpc, with scatter 30 Mpc. Similar results are obtained when excluding nuclear

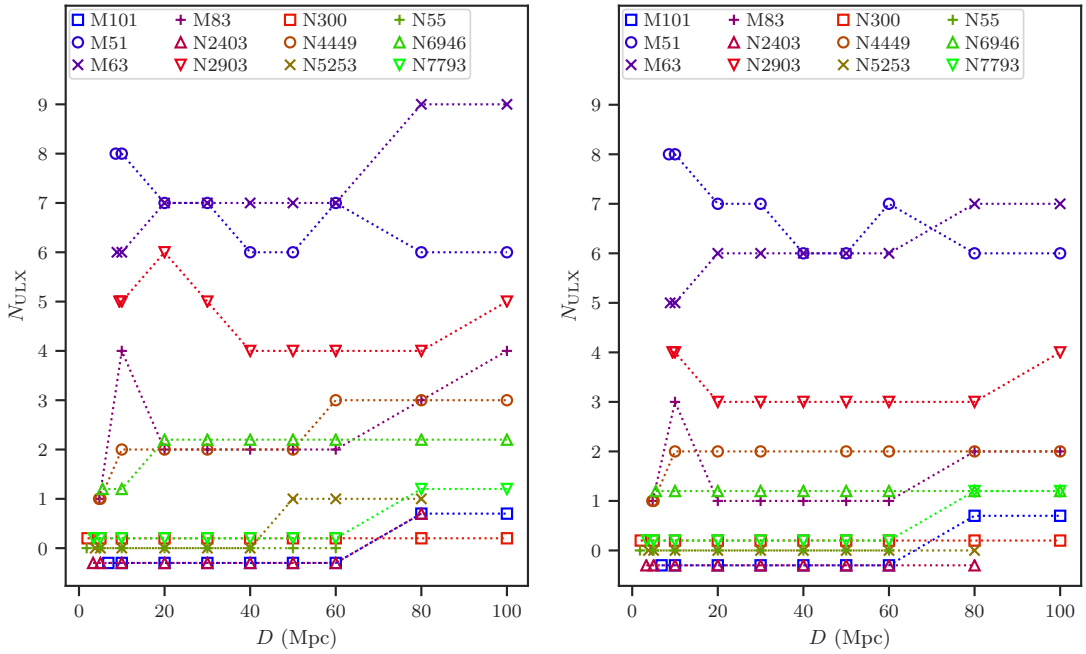


Figure 4.3.1: Left: the number of ULX candidates in each galaxy, as a function of the distance. Small (≤0.3) offsets have been applied in the data of a few galaxies, for visualisation purposes. **Right**: same as the left panel, but excluding nuclear sources.

sources (62±15 Mpc with scatter of 34 Mpc). Due to the small number of sources, this experiment cannot provide a statistically significant answer to our question.

4.3.3 Confusion distance results

Figure 4.3.2 shows the confusion distance estimates for the galaxies in our sample, evaluated at a grid of luminosity thresholds extending to the full range of X-ray luminosities in our sample*. We observe that the confusion distance increases with luminosity. Around the ULX limit, $L_X > 10^{39}$ erg s^{-1}, the confusion distance is observed at 40−100 Mpc.

To quantify the average trend in our sample, we apply a Kernel smoothing method to the d_C estimates of all galaxies, as function of luminosity. Specifically, we use local linear regression which accounts for changes in the slope in the data (e.g.,

* We selected a spacing of 0.01 dex for visualisation purposes. We note that this value does not affect the results. The metric d_C exhibits a 'cumulative logic', similar to the cumulative luminosity functions which can be evaluated at any interval.

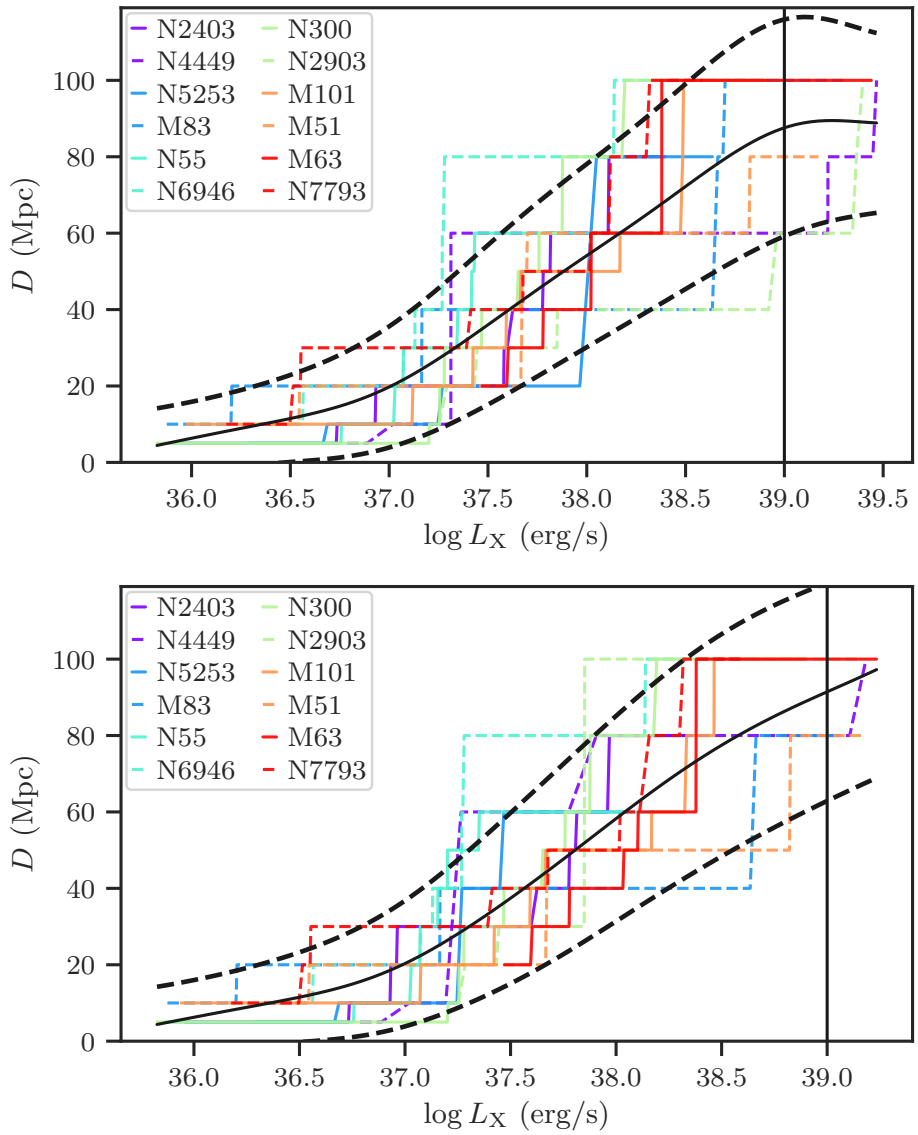


Figure 4.3.2: The confusion distance d_C for each galaxy (see labels), and for all (top) or off-nuclear sources (bottom). The black line represents the average confusion distance, obtained by smoothing the data points from all galaxies, whereas the dashed lines encompass the 68% CI of the average. At the ULX limit (10^{39} erg s^{-1}; see vertical line), the average confusion distance when accounting for all sources is 87±28 Mpc, whereas after excluding nuclear sources is 91±29 Mpc.

Cleveland 1979), with a Gaussian kernel of bandwidth equal to 0.5 dex in X-ray luminosity. Similarly, we calculate the local scatter at any given luminosity using the weighted standard deviation with the weights computed using the same kernel as above. The average confusion distance increases with luminosity, as expected from the estimates from individual galaxies, and presents a flattening above L_X=10^{39} erg s^{-1}. At L_X>10^{38} erg s^{-1}, and 10^{39} erg s^{-1}, we find d_C=54±24 Mpc, and 87±28 Mpc, respectively. Similar results are found when excluding nuclear sources: d_C(10^{38} erg s^{-1})=58±27 Mpc and d_C(10^{39} erg s^{-1})=91±29 Mpc.

4.4 Discussion

4.4.1 The trends of slopes with distance

The investigation of the decrease of the X-ray luminosity function power-law slope in §4.3.1 indicates that in more than half of the galaxies in our sample (seven), the slope decreases at distances larger than 40 Mpc.

In four galaxies, however, the slopes are consistent with being constant along the different distances, possibly due to the large uncertainties in the slope estimates, or the small source density: two of these galaxies are the ones with the lowest number of sources (NGC 4449 and NGC 5253).

Notably, the other two galaxies with almost constant slope, M51 and M63, along with the galaxy in which we observe an increase of the slope with distance, NGC 2903, are the most ULX-rich in our sample (at least four off-nuclear sources with X-ray luminosity >10^{39} erg s^{-1}). The presence of ULXs is indicative of a highly star-forming environment, the clumpy nature of which, may affect significantly the spatial scales at which source confusion is important. Instead, the lack of systematic trends could reflect differences in the spatial distribution between the ULXs and the fainter sources (i.e., bright sources, including ULXs are more sparsely distributed, resulting in lower confusion, while lower luminosity sources may be subject to confusion). In addition, this could be the result of the rejection of sources that appear as extended sources after binning the images. Such sources may not be recovered by the detection process because of their larger spatial extent

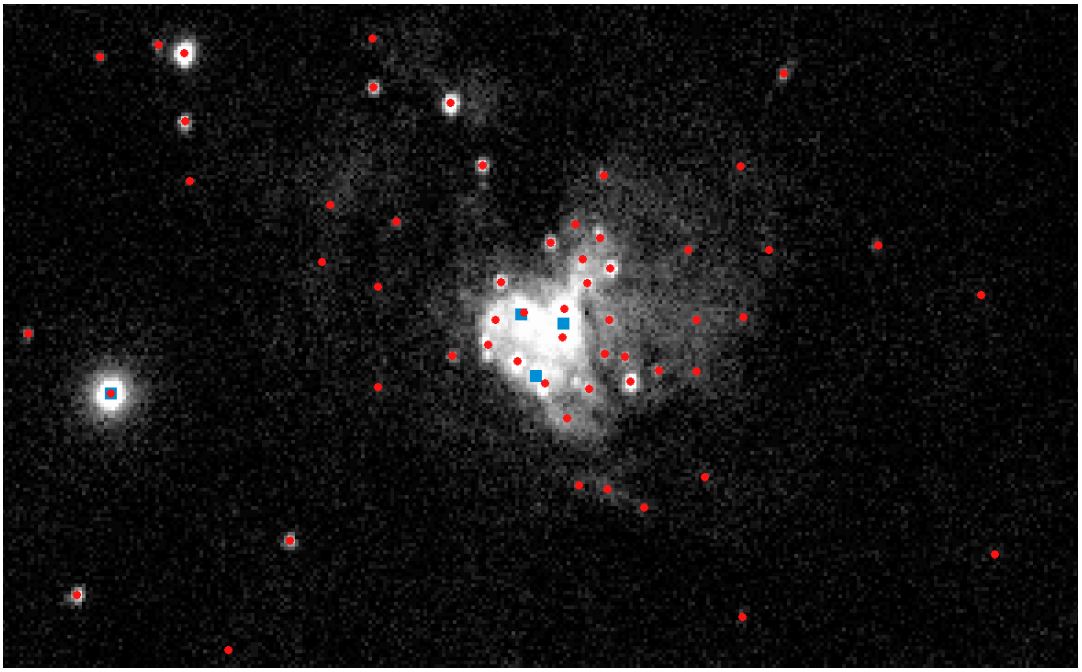


Figure 4.4.1: A 2.4×1.5 arcmin^2 region around the central part of M83 in the original X-ray image. Sources detected at the original distance are shown by red circles. Cyan squares denote ULXs detected at the binned image corresponding to a distance of 10 Mpc. The known ULX in M83 is detected at all distances (towards the left part of the image). The three additional ULX candidates close to the nucleus, appear at 10 Mpc, only to become blended into one nuclear source at $D \geq 20$ Mpc.

4.4.2 Number of ULXs

As discussed in §4.3.2 (see also Figure 4.3.1), the number of ULX as a function of fiducial distance is not a very reliable confusion metric.

The fluctuations in the number of ULXs as a function of the simulated distance, in three galaxies, can be attributed to the diffuse emission clumps. In the case of M83, at its original distance of 5 Mpc, we detect one source above 10^{39} erg s^{-1}, which is a known ULX (CXOU J133705.1-295207; Soria *et al.* 2012). However, at D=10 Mpc, in the crowded, and diffuse-emission rich, circumnuclear region of the host galaxy, we detect three additional ULX candidates. This is supported by the fact that the sums of the luminosities of all sources, in the original resolution (unbinned image), detected within a 5 arcsec radius from the confused sources, are less than the luminosities of the latter. At distances ≥20 Mpc, these three sources, are blended together, so that only two ULXs appear: the catalogued one, and the nuclear source. In Figure 4.4.1 we

show the above mentioned sources on the original X-ray image used for our analysis.

The same effect is observed in NGC 2903. Two out of the five ULX candidates at 10 Mpc, are found close to the nucleus. At D=20 Mpc, an a additional circum-nuclear source is detected. However, the three nuclear sources at these distances, are gradually blended into two and finally one nuclear source, at 30 and 40 Mpc respectively.

In the third such galaxy, M51, with an original distance of 9 Mpc, we detect eight ULX candidates. Two of them are found close to the nucleus, and become blended at 20 Mpc. Four are found at large separations from the centre (≈2 arcmin) and are detected as unique sources at all probed distances. However, the remaining two ULXs, are found also at large separations from the centre, but being close to each other (≈8arcsec), and are confused at $D \geq 40$ Mpc.

The above fluctuations reveal the significance of the source confusion effect with respect to the diffuse emission clumps, and the source-crowded regions, whether circum-nuclear or star-forming. However, the fluctuations observed at $D \lesssim 30$ Mpc may not affect significantly the statistics of large-scale demographic surveys, since they average out as observed in our sample. On the other hand, the increase in the number of ULXs observed at larger distances, is systematically observed in all galaxies, for $D > 60$ Mpc.

Another complication that becomes apparent from Figure 4.3.1 is the decrease in the number of ULXs for some galaxies and distances <20 Mpc. We attribute this to the rejection of extended clumps of emission by the source detection algorithm. In this case, ULXs associated or close to other sources or diffuse emission clumps may form extended structures when slightly binned. At larger distances these clumps will appear point-like again, and the sources will be recovered.

However, there are galaxies for which no ULX is observed at all probed distances (NGC 55, and NGC 300). This is expected due to their low total X-ray output, as suggested by the integrated luminosity of the sources at the original distance being close to the ULX luminosity threshold ($\log \sum L_X$ =39.2 and 38.8, respectively).

4.4.3 Confusion distance

The confusion distance, d_C, increases with luminosity in all the galaxies in our sample. This is attributed to the power-law nature of the XRBs: sources of low luminosities present high surface densities, hence they will show confusion even

at small distances. The flattening observed at high-luminosity end of the plots in Figure 4.3.2 is a manifestation of the lack of simulated distances above 100 Mpc and the sparsity of the luminous sources.

We find that around the ULX limit, L_X=10^{39} erg s^{-1}, the confusion distance is consistently above the fiducial limit of 40 Mpc adopted by us (see §3) and other surveys of HMXBs (e.g., Mineo *et al.* 2012). The average d_C is ~90±30 Mpc, regardless if we consider nuclear regions.

Appendix

Appendix 4.A Luminosity functions and slope plots

The LFs of all the X-ray sources in the D_{25} regions of the sample galaxies, and for all the distances considered (original and simulated) are shown in Figure 4.A.1. The corresponding LFs when excluding nuclear sources are shown in Figure 4.A.2. Similarly, the power-law slopes calculated in §4.2.3.1, are shown in Figure 4.A.3 (for all sources), and Figure 4.A.2 (for all off-nuclear sources).

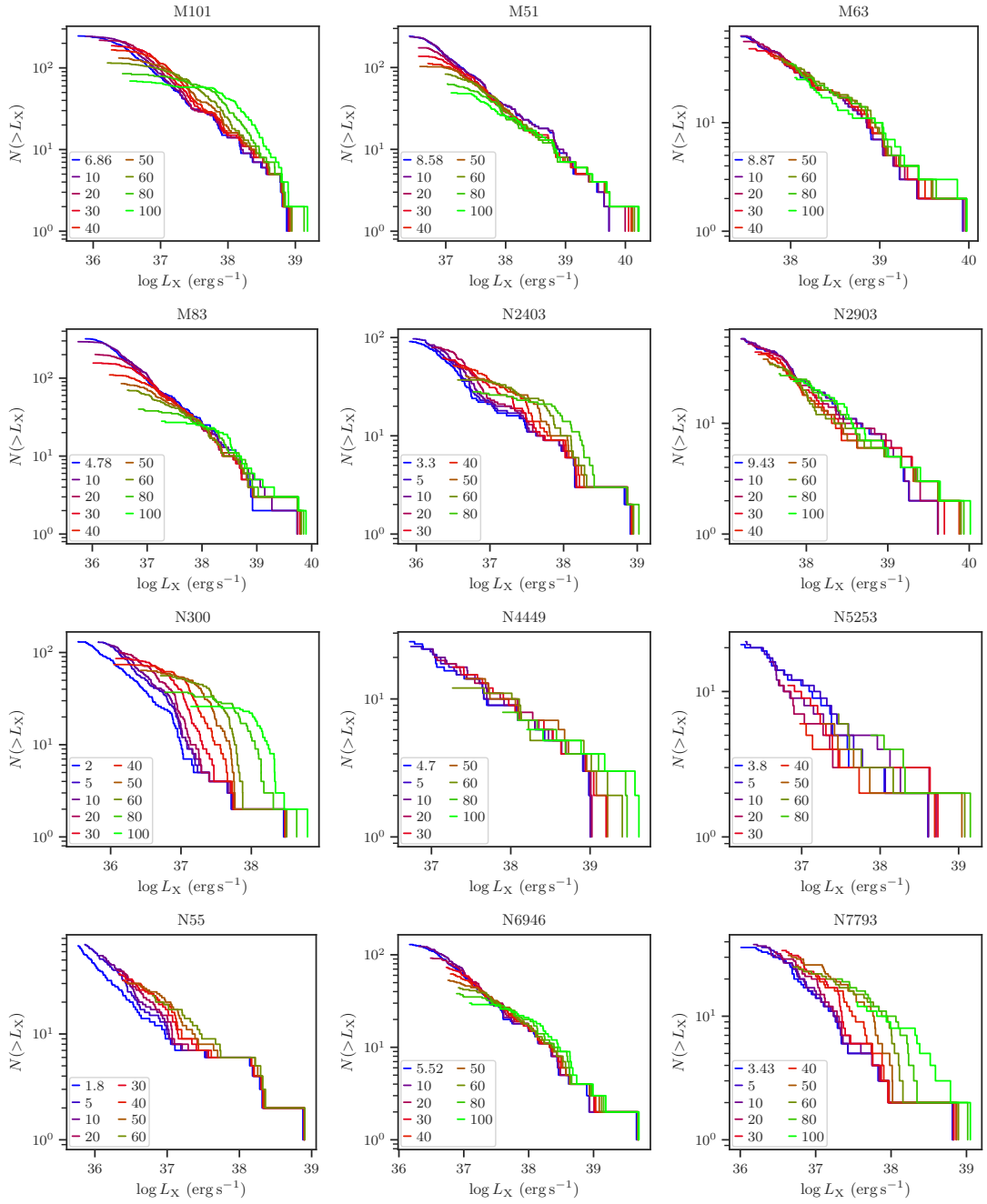


Figure 4.A.1: X-ray luminosity functions of all sources in the sample galaxies (plot titles). Colours denote different distances (original and simulated; see legend).

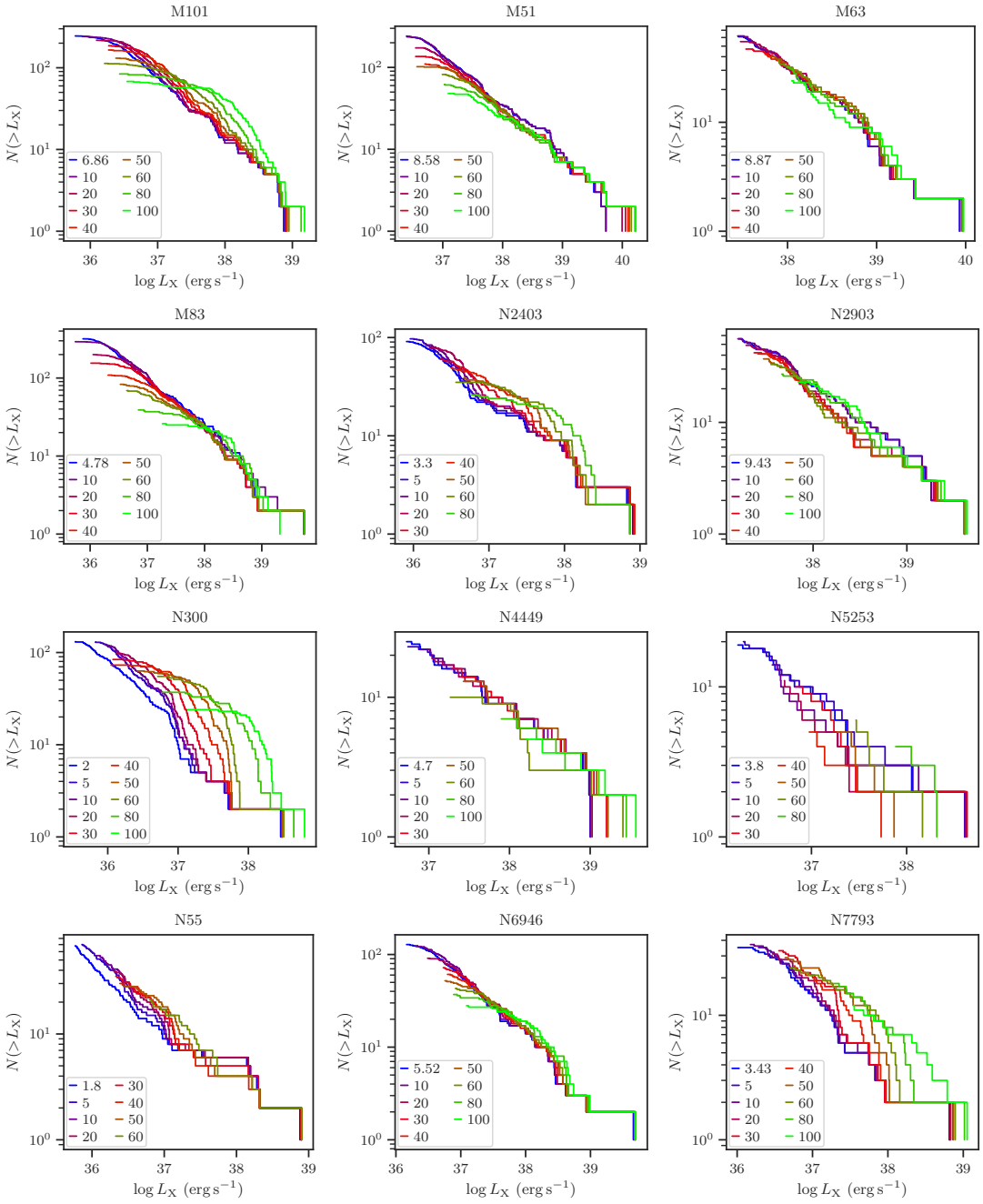


Figure 4.A.2: X-ray luminosity functions of off-nuclear sources in the sample galaxies (plot titles). Colours denote different distances (original and simulated; see legend).

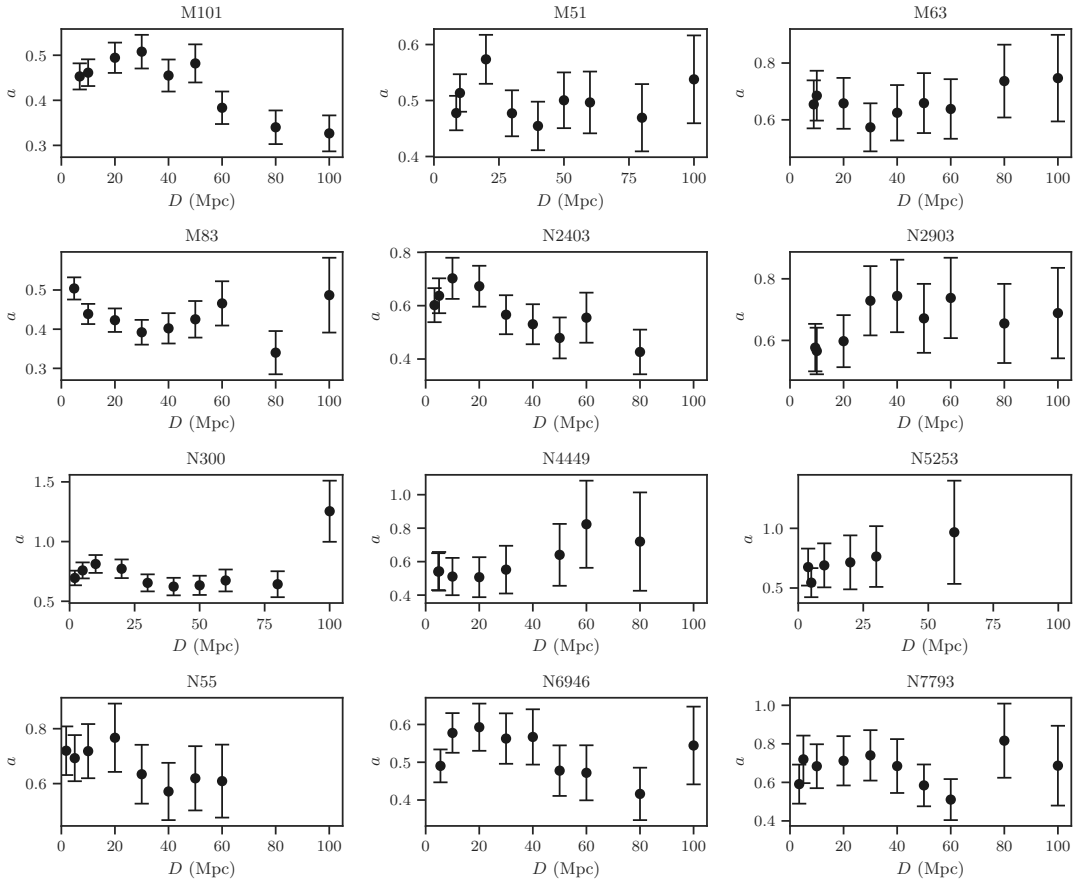


Figure 4.A.3: Slopes of the luminosity functions of all sources for each galaxy (plot titles), at the original and the simulated distances (x-axis). Missing points indicate distances at which we either do not have data, or the number of sources was less than five.

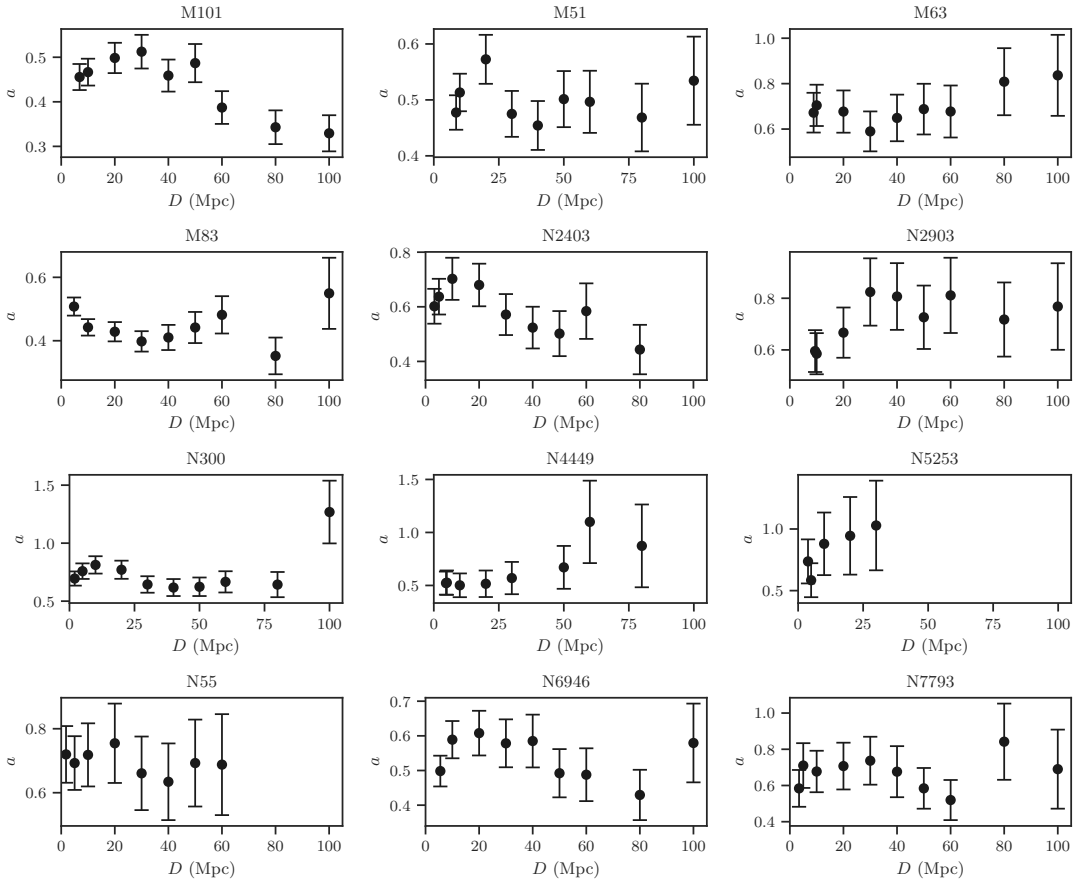


Figure 4.A.4: Slopes of the luminosity functions of off-nuclear sources for each galaxy (plot titles), at the original and the simulated distances (x-axis). Missing points indicate distances at which we either do not have data, or the number of sources was less than five.

This page intentionally left blank

5

Conclusions and future perspectives

In this thesis we presented: (i) a new value-added galaxy catalogue, the *HECATE*, (ii) a census of the ULXs in the local Universe based on the *HECATE* and *CSC 2.0*, and (iii) a study of the source confusion in the ULX regime. In the following sections we present a summary of the these results, as well as future perspectives.

5.1 The Heraklion Extragalactic Catalogue

We compile an all-sky galaxy catalogue providing the necessary information for a detailed study of ULX populations in the local Universe. For this reason, we:

1. Selected all known galaxies up to a distance of 200 Mpc from the *HyperLEDA* database, and obtained their positions and sizes.
2. Estimated redshift-independent distances using published measurements, and – in a consistent way – redshift-dependent distances for galaxies without measurement.

3. Associated the sample with different multi-wavelength surveys, ensuring that the photometry of nearby and extended galaxies accounts for their integrated flux.
4. We calculated SFRs, stellar masses, and metallicities using the multi-wavelength data.
5. Obtained morphological classifications from *HyperLEDA*.
6. Classified galaxies on the basis of their nuclear activity.

The careful treatment for nearby and extended galaxies (e.g., z-independent distances, extended photometry), the large distance limit (200 Mpc), the all-sky coverage, and its completeness in terms of the provided data, make the *HECATE* an ideal sample for a wide range of applications in Astrophysics. For this reason, the catalogue has been used for a study of the properties of the galaxies that will be detected by the *eROSITA* all-sky survey in the X-rays (Basu-Zych *et al.* 2020), censuses of ULX populations with *Chandra* (Kovlakas *et al.* 2020) and *XMM-Newton* data (Bernadich et al., in preparation), and X-ray scaling relations of galaxies (Anastasopoulou et al., in preparation). The *HECATE* can also be used in future studies for the prioritisation of galaxies for follow-up surveys of electromagnetic counterparts of gravitation-wave sources, and studies of the origin of cosmic-ray and gamma-ray sources. In addition, the *HECATE* can be used for the the initial characterisation of sources in current and upcoming all-sky and serendipitous surveys (e.g., *eROSITA*), as well as transient sources such as supernovæ and tidal disruption events (e.g., with *LSST*).

5.2 Ultraluminous X-ray source populations

The wealth of the information provided in the *HECATE* was crucial for addressing the main challenges in ULX demographic studies. The range and coverage of the *HECATE* in terms of the provided data, in combination with the *CSC 2.0*, allowed us to perform the most comprehensive study of ULX populations. In this way, we were able to compare our results with previous works, test the observed trends of ULX populations, and constrain the parameters of the scaling relations with an unprecedented accuracy. Specifically:

1. We find that the average scaling of the ULXs with SFR in late-type galaxies is 0.51 ± 0.06 ULXs per $M_{\odot}$ yr^{-1}, which is consistent with the the HMXB luminosity

function of Mineo *et al.* (2012), confirming the expectation that ULXs in spiral galaxies are predominantly HMXBs.

2. When accounting for sample differences, the average scaling of the ULXs with SFR agrees with previous estimates (e.g. Swartz *et al.* 2011).

3. The ULX-SFR scaling is a function of the morphological type of the host. We find that the significant increase in galaxies of Sc type or later, coincides with the lower metallicity in these galaxies, indicating the metallicity effect as the driver of the trend with the morphological type. We find an excess of ULXs in low-mass late-type galaxies, which we attributed to the metallicity effect.

4. For the first time, we present a ULX scaling relation with both SFR and stellar mass to account for the contribution of both LMXB and HMXB populations in the ULX regime. While on average the $M_{\star}$ dependence can be neglected, we find that when focusing on early-type spirals (S0/a-Sbc), it becomes significant. This scaling relation is consistent with recent XRB LFs accounting for both stellar population parameters (Lehmer *et al.* 2019).

5. The scaling of ULXs with stellar mass, i.e. the specific ULX frequency, in early-type galaxies (6.3 ± 1.0 ULXs per $10^{12}\mathrm{M}_{\odot}$) agrees with the LMXB LF (Zhang *et al.* 2012), and previous demographic studies. However, this agreement is achieved only when comparing estimates based on galaxy samples of similar stellar masses. For the first time, we find that the specific ULX frequency decreases when stellar mass increases.

6. The aforementioned evolution of specific ULX frequency with the stellar mass, is attributed to the star-formation history differences of early-type galaxies of different masses. We support this by showing the same trend in the convolution of empirical SFHs of early-type galaxies (McDermid *et al.* 2015) with binary population synthesis models (Wiktorowicz *et al.* 2017).

7. We find evidence for a small but significant ULX population in early-type galaxies at high galactocentric distances. This population may be due to ULXs residing in globular clusters.

5.3 Source confusion

Using X-ray images from the *Chandra* archive, for twelve nearby (D<10 Mpc) galaxies, we simulated their observations at distances in the range 10–100 Mpc. This way, we studied the effect of source confusion on the luminosity functions of the detect sources, and the number of sources above the ULX limit, 10^{39} erg s^{-1}. We found the following:

1. The slope of the luminosity function decreases with distance in the majority of the galaxies in our sample, especially when comparing simulations for distances below and above the commonly-used limit of 40 Mpc for *Chandra* surveys.
2. The ULX-rich galaxies do not show any significant change in their LF slopes with distance, which may be the result of: (a) different star-formation conditions affecting the spatial scales at which source confusion is important, (b) differences between the spatial distributions of ULXs and fainter sources, (c) blended sources which may appear as extended, and therefore are excluded by the detection process which focuses point-like sources.
3. Sources at the high surface density nuclei, and diffuse emission clumps, may appear as ULXs even at small distances (∼20 Mpc), and are blended into single nuclear sources at large distances.
4. While the fluctuations in the observed number of ULXs in individual galaxies at $D \leq 40$ Mpc seem to average out in large samples, at larger distances we see a systematic bias in the direction of more luminous sources.
5. A newly proposed metric, the *confusion distance*, taking into account the entire luminosity distributions of the sources, indicates that the distance at which ULX populations are significantly affected by source confusion, is 90±30 Mpc. This limit is much lower for the more densely distributed lower-luminosity sources.

5.4 Future perspectives

The ULX population in a galaxy is the result of its initial binary parameters, and the star-formation and metallicity history, the evolution pathways of ULXs, and the physical processes responsible for their accretion and X-ray emission. Our published census of ULXs (Kovlakas *et al.* 2020), thanks to its size and the diversity in the galactic environments, provides a unique opportunity to compare observations with theoretical

models. By combining galaxies from cosmological simulations, and the forthcoming state-of-the-art binary population synthesis code *POSYDON* (Fragos et al. 2021), we will constrain the parameters of ULX formation and evolution models, such as the supernova kick-velocity distribution, initial mass ratio, common envelope ejection efficiency, mass-loss rate, etc.) These new constrains will also provide useful input for the formation and evolution of massive binaries in general, including progenitors of gravitational-wave sources and short gamma-ray bursts.

One major constraint of studies of the scaling relations and luminosity functions of XRBs, is the imposition of a distance limit to avoid source confusion. This, unavoidably reduces the accuracy of the results due to the smaller sample sizes, but most importantly, it limits the diversity of the galaxies explored. Our exploration lays the groundwork for a future detailed study of the source blending effect in the X-rays sources at larger galaxy distances. Using observations and semi-analytical models, we will provide the methodology for correcting the source confusion effect in scaling relations and LFs, extending their range in terms of the distance and the stellar population parameter space, without the need for additional observational campaigns. This will also be important for ULX studies at higher redshifts (such as those detected in deep surveys), or with lower spatial resolution telescopes (e.g., *XMM-Newton*, *eROSITA*).

The *HECATE* is a continuing project: future development of the *HECATE* will extend its volume beyond the current redshift limit of z=0.047 (corresponding to $\approx$200 Mpc), and address its limitations in terms of the data coverage. By: (i) incorporating additional multi-wavelength data, (ii) using forced photometry techniques for wavebands for which there is no available extended source photometry, and (iii) applying SED analysis, we will obtain additional estimates of the stellar population parameters, for a larger fraction of the galaxies in the *HECATE*. These improvements will widen the range of applications of the catalogue in X-ray astronomy, and extragalactic astrophysics in general.

This page intentionally left blank

Bibliography

Abbott, B.P. *et al.* (2017a). A gravitational-wave standard siren measurement of the Hubble constant. Nature, **551**, 85–88. (in p. 43)

Abbott, B.P. *et al.* (2017b). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett., **119**, 161101. (in pp. 10, 16, and 43)

Abbott, B.P. *et al.* (2017c). On the Progenitor of Binary Neutron Star Merger GW170817. ApJ, **850**, L40. (in pp. 43 and 48)

Abbott, B.P. *et al.* (2020). Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, **23**, 3. (in p. 43)

Abramowicz, M.A., Czerny, B., Lasota, J.P. & Szuszkiewicz, E. (1988). Slim Accretion Disks. ApJ, **332**, 646. (in pp. 10 and 64)

Ackermann, M., Ajello, M., Allafort, A., Baldini, L., Ballet, J., Bastieri, D., Bechtol, K., Bellazzini, R., Berenji, B., Bloom, E.D., Bonamente, E., Borgland , A.W., Bouvier, A., Bregeon, J., Brigida, M., Bruel, P., Buehler, R., Buson, S., Caliandro, G.A., Cameron, R.A., Caraveo, P.A., Casandjian, J.M., Cecchi, C., Charles, E., Chekhtman, A., Cheung, C.C., Chiang, J., Cillis, A.N., Ciprini, S., Claus, R., Cohen-Tanugi, J., Conrad, J., Cutini, S., de Palma, F., Dermer, C.D., Digel, S.W., Silva, E.d.C.e., Drell, P.S., Drlica-Wagner, A., Favuzzi, C., Fegan, S.J., Fortin, P., Fukazawa, Y., Funk, S., Fusco, P., Gargano, F., Gasparrini, D., Germani, S., Giglietto, N., Giordano, F., Glanzman, T., Godfrey, G., Grenier, I.A., Guiriec, S., Gustafsson, M., Hadasch, D., Hayashida, M., Hays, E., Hughes, R.E., Jóhannesson, G., Johnson, A.S., Kamae, T., Katagiri, H., Kataoka, J., Knödlseder, J., Kuss, M., Lande, J., Longo, F., Loparco, F., Lott, B., Lovellette, M.N., Lubrano, P., Madejski, G.M., Martin, P., Mazziotta, M.N., McEnery, J.E., Michelson, P.F., Mizuno, T., Monte, C., Monzani, M.E., Morselli, A., Moskalenko, I.V., Murgia, S., Nishino, S., Norris, J.P., Nuss, E., Ohno, M., Ohsugi, T., Okumura, A., Omodei, N., Orlando, E., Ozaki, M., Parent, D., Persic, M., Pesce-Rollins, M., Petrosian, V., Pierbattista, M., Piron, F., Pivato, G., Porter, T.A., Rainò, S., Rando, R., Razzano, M., Reimer, A., Reimer, O., Ritz, S., Roth, M., Sbarra, C., Sgrò, C., Siskind, E.J., Spandre, G., Spinelli, P., Stawarz, Ł., Strong, A.W., Takahashi, H., Tanaka, T., Thayer, J.B., Tibaldo, L., Tinivella, M., Torres, D.F., Tosti, G., Troja, E., Uchiyama, Y., Vandenbroucke, J., Vianello, G., Vitale, V., Waite, A.P., Wood, M. & Yang, Z. (2012). GeV Observations of Star-forming Galaxies with the Fermi Large Area Telescope. ApJ, **755**, 164. (in p. 16)

Bibliography

Adhikari, S., Fishbach, M., Holz, D.E., Wechsler, R.H. & Fang, Z. (2020). The binary-host connection: astrophysics of gravitational wave binaries from their host galaxy properties. *arXiv e-prints*, arXiv:2001.01025. (in p. 43)

Aguado, D.S., Ahumada, R., Almeida, A., Anderson, S.F., Andrews, B.H., Anguiano, B., Aquino Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., Aubert, M., Avila-Reese, V., Badenes, C., Barboza Rembold, S.r., Barger, K., Barrera-Ballesteros, J., Bates, D., Bautista, J., Beaton, R.L., Beers, T.C., Belfiore, F., Bernardi, M., Bershady, M., Beutler, F., Bird, J., Bizyaev, D., Blanc, G.A., Blanton, M.R., Blomqvist, M., Bolton, A.S., Boquien, M., Borissova, J., Bovy, J., Brand t, W.N., Brinkmann, J., Brownstein, J.R., Bundy, K., Burgasser, A., Byler, N., Cano Diaz, M., Cappellari, M., Carrera, R., Cervantes Sodi, B., Chen, Y., Cherinka, B., Choi, P.D., Chung, H., Coffey, D., Comerford, J.M., Comparat, J., Covey, K., da Silva Ilha, G., da Costa, L., Dai, Y.S., Damke, G., Darling, J., Davies, R., Dawson, K., de Sainte Agathe, V., Deconto Machado, A., Del Moro, A., De Lee, N., Diamond-Stanic, A.M., Domínguez Sánchez, H., Donor, J., Drory, N., du Mas des Bourboux, H., Duckworth, C., Dwelly, T., Ebelke, G., Emsellem, E., Escoffier, S., Fernández-Trincado, J.G., Feuillet, D., Fischer, J.L., Fleming, S.W., Fraser-McKelvie, A., Freischlad, G., Frinchaboy, P.M., Fu, H., Galbany, L., Garcia-Dias, R., García-Hernández, D.A., Garma Oehmichen, L.A., Geimba Maia, M.A., Gil-Marín, H., Grabowski, K., Gu, M., Guo, H., Ha, J., Harrington, E., Hasselquist, S., Hayes, C.R., Hearty, F., Hernandez Toledo, H., Hicks, H., Hogg, D.W., Holley-Bockelmann, K., Holtzman, J.A., Hsieh, B.C., Hunt, J.A.S., Hwang, H.S., Ibarra-Medel, H.J., Jimenez Angel, C.E., Johnson, J., Jones, A., Jönsson, H., Kinemuchi, K., Kollmeier, J., Krawczyk, C., Kreckel, K., Kruk, S., Lacerna, I., Lan, T.W., Lane, R.R., Law, D.R., Lee, Y.B., Li, C., Lian, J., Lin, L., Lin, Y.T., Lintott, C., Long, D., Longa-Peña, P., Mackereth, J.T., de la Macorra, A., Majewski, S.R., Malanushenko, O., Manchado, A., Maraston, C., Mariappan, V., Marinelli, M., Marques-Chaves, R., Masseron, T., Masters, K.L., McDermid, R.M., Medina Peña, N., Meneses-Goytia, S., Merloni, A., Merrifield, M., Meszaros, S., Minniti, D., Minsley, R., Muna, D., Myers, A.D., Nair, P., Correa do Nascimento, J., Newman, J.A., Nitschelm, C., Olmstead, M.D., Oravetz, A., Oravetz, D., Ortega Minakata, R.A., Pace, Z., Padilla, N., Palicio, P.A., Pan, K., Pan, H.A., Parikh, T., Parker, I., James, Peirani, S., Penny, S., Percival, W.J., Perez-Fournon, I., Peterken, T., Pinsonneault, M.H., Prakash, A., Raddick, M.J., Raichoor, A., Riffel, R.A., Riffel, R., Rix, H.W., Robin, A.C., Roman-Lopes, A., Rose, B., Ross, A.J., Rossi, G., Rowlands, K., Rubin, K.H.R., Sánchez, S.F., Sánchez-Gallego, J.R., Sayres, C., Schaefer, A., Schiavon, R.P., Schimoia, J.S., Schlafly, E., Schlegel, D., Schneider, D.P., Schultheis, M., Seo, H.J., Shamsi, S.J., Shao, Z., Shen, S., Shetty, S., Simonian, G., Smethurst, R.J., Sobeck, J., Souter, B.J., Spindler, A., Stark, D.V., Stassun, K.G., Steinmetz, M., Storchi-Bergmann, T., Stringfellow, G.S., Suárez, G., Sun, J., Taghizadeh-Popp, M., Talbot, M.S., Tayar, J., Thakar, A.R., Thomas, D., Tissera, P., Tojeiro, R., Troup, N.W., Unda-Sanzana, E., Valenzuela, O., Vargas-Magaña, M., Vázquez-Mata, J.A., Wake, D., Weaver, B.A., Weijmans, A.M., Westfall, K.B., Wild, V., Wilson, J., Woods, E., Yan, R., Yang, M., Zamora, O., Zasowski, G., Zhang, K., Zheng, Z., Zheng, Z., Zhu,

G., Zinn, J.C. & Zou, H. (2019). The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stellar Library. ApJS, **240**, 23. (in p. 56)

Anastasopoulou, K., Zezas, A., Ballo, L. & Della Ceca, R. (2016). A deep Chandra observation of the interacting star-forming galaxy Arp 299. MNRAS, **460**, 3570–3586. (in pp. 11, 65, and 73)

Anastasopoulou, K., Zezas, A., Gkiokas, V. & Kovlakas, K. (2019). Do sub-galactic regions follow the galaxy-wide X-ray scaling relations? The example of NGC 3310 and NGC 2276. MNRAS, **483**, 711–733. (in p. 97)

Andrews, J.J. & Zezas, A. (2019). Double neutron star formation: merger times, systemic velocities, and travel distances. MNRAS, **486**, 3213–3227. (in p. 48)

Angelini, L., Loewenstein, M. & Mushotzky, R.F. (2001). The X-Ray Globular Cluster Population in NGC 1399. ApJ, **557**, L35–L38. (in pp. 11, 64, and 65)

Antoniou, V. & Zezas, A. (2016). Star formation history and X-ray binary populations: the case of the Large Magellanic Cloud. MNRAS, **459**, 528–553. (in pp. 8 and 101)

Antoniou, V., Zezas, A., Drake, J.J., Badenes, C., Haberl, F., Wright, N.J., Hong, J., Di Stefano, R., Gaetz, T.J., Long, K.S., Plucinsky, P.P., Sasaki, M., Williams, B.F., Winkler, P.F. & SMC XVP collaboration (2019). Deep Chandra Survey of the Small Magellanic Cloud. III. Formation Efficiency of High-mass X-Ray Binaries. ApJ, **887**, 20. (in p. 101)

Arcavi, I., McCully, C., Hosseinzadeh, G., Howell, D.A., Vasylyev, S., Poznanski, D., Zaltzman, M., Maoz, D., Singer, L., Valenti, S., Kasen, D., Barnes, J., Piran, T. & Fong, W.f. (2017). Optical Follow-up of Gravitational-wave Events with Las Cumbres Observatory. ApJ, **848**, L33. (in pp. 43 and 44)

Artale, M.C., Mapelli, M., Giacobbo, N., Sabha, N.B., Spera, M., Santoliquido, F. & Bressan, A. (2019). Host galaxies of merging compact objects: mass, star formation rate, metallicity, and colours. MNRAS, **487**, 1675–1688. (in pp. 16 and 43)

Artale, M.C., Bouffanais, Y., Mapelli, M., Giacobbo, N., Sabha, N.B., Santoliquido, F., Pasquato, M. & Spera, M. (2020a). An astrophysically motivated ranking criterion for low-latency electromagnetic follow-up of gravitational wave events. MNRAS, **495**, 1841–1852. (in p. 45)

Artale, M.C., Mapelli, M., Bouffanais, Y., Giacobbo, N., Pasquato, M. & Spera, M. (2020b). Mass and star formation rate of the host galaxies of compact binary mergers across cosmic time. MNRAS, **491**, 3419–3434. (in pp. 43, 44, and 47)

Ashby, M.L.N., Mahajan, S., Smith, H.A., Willner, S.P., Fazio, G.G., Raychaudhury, S., Zezas, A., Barmby, P., Bonfini, P., Cao, C., González-Alfonso, E., Ishihara, D., Kaneda, H., Lyttle, V., Madden, S., Papovich, C., Sturm, E., Surace, J., Wu, H. & Zhu, Y.N. (2011). The Star Formation Reference Survey. I. Survey Description and Basic Data. PASP, **123**, 1011. (in p. 110)

Bibliography

Bachetti, M., Harrison, F.A., Walton, D.J., Grefenstette, B.W., Chakrabarty, D., Fürst, F., Barret, D., Beloborodov, A., Boggs, S.E., Christensen, F.E., Craig, W.W., Fabian, A.C., Hailey, C.J., Hornschemeier, A., Kaspi, V., Kulkarni, S.R., Maccarone, T., Miller, J.M., Rana, V., Stern, D., Tendulkar, S.P., Tomsick, J., Webb, N.A. & Zhang, W.W. (2014). An ultraluminous X-ray source powered by an accreting neutron star. Nature, **514**, 202–204. (in pp. 9 and 64)

Basu-Zych, A.R., Lehmer, B.D., Hornschemeier, A.E., Bouwens, R.J., Fragos, T., Oesch, P.A., Belczynski, K., Brandt, W.N., Kalogera, V., Luo, B., Miller, N., Mullaney, J.R., Tzanavaris, P., Xue, Y. & Zezas, A. (2013a). The X-Ray Star Formation Story as Told by Lyman Break Galaxies in the 4 Ms CDF-S. ApJ, **762**, 45. (in pp. 8, 10, 65, and 97)

Basu-Zych, A.R., Lehmer, B.D., Hornschemeier, A.E., Gonçalves, T.S., Fragos, T., Heckman, T.M., Overzier, R.A., Ptak, A.F. & Schiminovich, D. (2013b). Evidence for Elevated X-Ray Emission in Local Lyman Break Galaxy Analogs. ApJ, **774**, 152. (in pp. 8, 10, 65, and 97)

Basu-Zych, A.R., Lehmer, B., Fragos, T., Hornschemeier, A., Yukita, M., Zezas, A. & Ptak, A. (2016). Exploring the Overabundance of ULXs in Metal- and Dust-poor Local Lyman Break Analogs. ApJ, **818**, 140. (in pp. 65, 73, 97, and 117)

Basu-Zych, A.R., Hornschemeier, A.E., Haberl, F., Vulic, N., Wilms, J., Zezas, A., Kovlakas, K., Ptak, A. & Dauser, T. (2020). The Next Generation X-ray Galaxy Survey with eROSITA. *arXiv e-prints*, arXiv:2008.01870. (in pp. 42, 104, and 136)

Bavera, S.S., Fragos, T., Qin, Y., Zapartas, E., Neijssel, C.J., Mandel, I., Batta, A., Gaebel, S.M., Kimball, C. & Stevenson, S. (2020). The origin of spin in binary black holes. Predicting the distributions of the main observables of Advanced LIGO. A&A, **635**, A97. (in p. 43)

Begelman, M.C. (1978). Black holes in radiation-dominated gas: an analogue of the Bondi accretion problem. MNRAS, **184**, 53–67. (in p. 10)

Begelman, M.C. (2002). Super-Eddington Fluxes from Thin Accretion Disks? ApJ, **568**, L97–L100. (in pp. 9 and 64)

Belczynski, K., Bulik, T., Fryer, C.L., Ruiter, A., Valsecchi, F., Vink, J.S. & Hurley, J.R. (2010). On the Maximum Mass of Stellar Black Holes. ApJ, **714**, 1217–1226. (in pp. 12, 65, and 95)

Bell, E.F., McIntosh, D.H., Katz, N. & Weinberg, M.D. (2003). The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. ApJS, **149**, 289–312. (in pp. 28, 31, 39, and 69)

Berger, E. (2014). Short-Duration Gamma-Ray Bursts. ARA&A, **52**, 43–105. (in p. 64)

Bhattacharjee, P. (2000). Origin and propagation of extremely high energy cosmic rays. Phys. Rep., **327**, 109–247. (in p. 49)

Blanton, M.R. & Roweis, S. (2007). K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared. AJ, **133**, 734–754. (in p. 31)

Brandt, W.N. & Alexander, D.M. (2015). Cosmic X-ray surveys of distant active galaxies. The demographics, physics, and ecology of growing supermassive black holes. A&A Rev., **23**, 1. (in p. 80)

Brinchmann, J., Charlot, S., White, S.D.M., Tremonti, C., Kauffmann, G., Heckman, T. & Brinkmann, J. (2004). The physical properties of star-forming galaxies in the low-redshift Universe. MNRAS, **351**, 1151–1179. (in pp. 15, 30, 35, and 69)

Brorby, M., Kaaret, P. & Prestwich, A. (2014). X-ray binary formation in low-metallicity blue compact dwarf galaxies. MNRAS, **441**, 2346–2353. (in pp. 8, 11, 65, 95, and 116)

Brorby, M., Kaaret, P., Prestwich, A. & Mirabel, I.F. (2016). Enhanced X-ray emission from Lyman break analogues and a possible L_X-SFR-metallicity plane. MNRAS, **457**, 4081–4088. (in pp. 10, 65, and 97)

Buikema, A., Cahillane, C., Mansell, G.L., Blair, C.D., Abbott, R., Adams, C., Adhikari, R.X., Ananyeva, A., Appert, S., Arai, K., Areeda, J.S., Asali, Y., Aston, S.M., Austin, C., Baer, A.M., Ball, M., Ballmer, S.W., Banagiri, S., Barker, D., Barsotti, L., Bartlett, J., Berger, B.K., Betzwieser, J., Bhattacharjee, D., Billingsley, G., Biscans, S., Blair, R.M., Bode, N., Booker, P., Bork, R., Bramley, A., Brooks, A.F., Brown, D.D., Cannon, K.C., Chen, X., Ciobanu, A.A., Clara, F., Cooper, S.J., Corley, K.R., Countryman, S.T., Covas, P.B., Coyne, D.C., Datrier, L.E.H., Davis, D., Di Fronzo, C., Dooley, K.L., Driggers, J.C., Dupej, P., Dwyer, S.E., Effler, A., Etzel, T., Evans, M., Evans, T.M., Feicht, J., Fernandez-Galiana, A., Fritschel, P., Frolov, V.V., Fulda, P., Fyffe, M., Giaime, J.A., Giardina, K.D., Godwin, P., Goetz, E., Gras, S., Gray, C., Gray, R., Green, A.C., Gustafson, E.K., Gustafson, R., Hanks, J., Hanson, J., Hardwick, T., Hasskew, R.K., Heintze, M.C., Helmling-Cornell, A.F., Holland, N.A., Jones, J.D., Kandhasamy, S., Karki, S., Kasprzack, M., Kawabe, K., Kijbunchoo, N., King, P.J., Kissel, J.S., Kumar, R., Landry, M., Lane, B.B., Lantz, B., Laxen, M., Lecoeuche, Y.K., Leviton, J., Liu, J., Lormand, M., Lundgren, A.P., Macas, R., MacInnis, M., Macleod, D.M., Márka, S., Márka, Z., Martynov, D.V., Mason, K., Massinger, T.J., Matichard, F., Mavalvala, N., McCarthy, R., McClelland, D.E., McCormick, S., McCuller, L., McIver, J., McRae, T., Mendell, G., Merfeld, K., Merilh, E.L., Meylahn, F., Mistry, T., Mittleman, R., Moreno, G., Mow-Lowry, C.M., Mozzon, S., Mullavey, A., Nelson, T.J.N., Nguyen, P., Nuttall, L.K., Oberling, J., Oram, R.J., O'Reilly, B., Osthelder, C., Ottaway, D.J., Overmier, H., Palamos, J.R., Parker, W., Payne, E., Pele, A., Penhorwood, R., Perez, C.J., Pirello, M., Radkins, H., Ramirez, K.E., Richardson, J.W., Riles, K., Robertson, N.A., Rollins, J.G., Romel, C.L., Romie, J.H., Ross, M.P., Ryan, K., Sadecki, T., Sanchez, E.J., Sanchez, L.E., Saravanan, T.R., Savage, R.L., Schaetzl, D., Schnabel, R., Schofield, R.M.S., Schwartz, E., Sellers, D., Shaffer, T., Sigg, D., Slagmolen, B.J.J., Smith, J.R., Soni, S., Sorazu, B., Spencer, A.P., Strain, K.A., Sun, L., Szczepańczyk, M.J., Thomas, M., Thomas, P., Thorne, K.A., Toland, K., Torrie, C.I., Traylor, G., Tse, M., Urban, A.L., Vajente, G., Valdes, G., Vand er-Hyde, D.C., Veitch, P.J., Venkateswara, K., Venugopalan, G., Viets, A.D., Vo, T., Vorvick, C., Wade, M., Ward, R.L., Warner, J., Weaver, B., Weiss, R., Whittle, C., Willke, B., Wipf, C.C., Xiao, L., Yamamoto, H., Yu, H., Yu, H., Zhang, L., Zucker, M.E. & Zweizig, J. (2020). Sensitivity and performance of the Advanced LIGO detectors in the third observing run. Phys. Rev. D, **102**, 062003. (in p. 43)

Carpano, S., Haberl, F., Maitra, C. & Vasilopoulos, G. (2018). Discovery of pulsations from NGC 300 ULX1 and its fast period evolution. MNRAS, **476**, L45–L49. (in pp. 9 and 64)

Chambers, K.C., Magnier, E.A., Metcalfe, N., Flewelling, H.A., Huber, M.E., Waters, C.Z., Denneau, L., Draper, P.W., Farrow, D., Finkbeiner, D.P., Holmberg, C., Koppenhoefer, J., Price, P.A., Rest, A., Saglia, R.P., Schlafly, E.F., Smartt, S.J., Sweeney, W., Wainscoat, R.J., Burgett, W.S., Chastel, S., Grav, T., Heasley, J.N., Hodapp, K.W., Jedicke, R., Kaiser, N., Kudritzki, R.P., Luppino, G.A., Lupton, R.H., Monet, D.G., Morgan, J.S., Onaka, P.M., Shiao, B., Stubbs, C.W., Tonry, J.L., White, R., Bañados, E., Bell, E.F., Bender, R., Bernard, E.J., Boegner, M., Boffi, F., Botticella, M.T., Calamida, A., Casertano, S., Chen, W.P., Chen, X., Cole, S., Deacon, N., Frenk, C., Fitzsimmons, A., Gezari, S., Gibbs, V., Goessl, C., Goggia, T., Gourgue, R., Goldman, B., Grant, P., Grebel, E.K., Hambly, N.C., Hasinger, G., Heavens, A.F., Heckman, T.M., Henderson, R., Henning, T., Holman, M., Hopp, U., Ip, W.H., Isani, S., Jackson, M., Keyes, C.D., Koekemoer, A.M., Kotak, R., Le, D., Liska, D., Long, K.S., Lucey, J.R., Liu, M., Martin, N.F., Masci, G., McLean, B., Mindel, E., Misra, P., Morganson, E., Murphy, D.N.A., Obaika, A., Narayan, G., Nieto-Santisteban, M.A., Norberg, P., Peacock, J.A., Pier, E.A., Postman, M., Primak, N., Rae, C., Rai, A., Riess, A., Riffeser, A., Rix, H.W., Röser, S., Russel, R., Rutz, L., Schilbach, E., Schultz, A.S.B., Scolnic, D., Strolger, L., Szalay, A., Seitz, S., Small, E., Smith, K.W., Soderblom, D.R., Taylor, P., Thomson, R., Taylor, A.N., Thakar, A.R., Thiel, J., Thilker, D., Unger, D., Urata, Y., Valenti, J., Wagner, J., Walder, T., Walter, F., Watters, S.P., Werner, S., Wood-Vasey, W.M. & Wyse, R. (2016). The Pan-STARRS1 Surveys. *arXiv e-prints*, arXiv:1612.05560. (in p. 42)

Chen, H.Y., Fishbach, M. & Holz, D.E. (2018). A two per cent Hubble constant measurement from standard sirens within five years. Nature, **562**, 545–547. (in p. 43)

Clark, G.W. (1975). X-ray binaries in globular clusters. ApJ, **199**, L143–L145. (in p. 5)

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836. (in p. 125)

Cluver, M.E., Jarrett, T.H., Dale, D.A., Smith, J.D.T., August, T. & Brown, M.J.I. (2017). Calibrating Star Formation in WISE Using Total Infrared Luminosity. ApJ, **850**, 68. (in pp. 33, 35, and 69)

Colbert, E.J.M. & Mushotzky, R.F. (1999). The Nature of Accreting Black Holes in Nearby Galaxy Nuclei. ApJ, **519**, 89–107. (in pp. 9 and 63)

Colbert, E.J.M. & Ptak, A.F. (2002). A Catalog of Candidate Intermediate-Luminosity X-Ray Objects. ApJS, **143**, 25–45. (in pp. 11, 64, and 65)

Cook, D.O., Kasliwal, M.M., Van Sistine, A., Kaplan, D.L., Sutter, J.S., Kupfer, T., Shupe, D.L., Laher, R.R., Masci, F.J., Dale, D.A., Sesar, B., Brady, P.R., Yan, L., Ofek, E.O., Reitze, D.H. & Kulkarni, S.R. (2019). Census of the Local Universe (CLU) Narrowband Survey. I. Galaxy Catalogs from Preliminary Fields. ApJ, **880**, 7. (in pp. 16, 37, and 43)

Crawford, D.F., Jauncey, D.L. & Murdoch, H.S. (1970). Maximum-Likelihood Estimation of the Slope from Number-Flux Counts of Radio Sources. ApJ, **162**, 405. (in p. 120)

Cutri, R.M., Skrutskie, M.F., van Dyk, S., Beichman, C.A., Carpenter, J.M., Chester, T., Cambresy, L., Evans, T., Fowler, J., Gizis, J., Howard, E., Huchra, J., Jarrett, T., Kopan, E.L., Kirkpatrick, J.D., Light, R.M., Marsh, K.A., McCallon, H., Schneider, S., Stiening, R., Sykes, M., Weinberg, M., Wheaton, W.A., Wheelock, S. & Zacharias, N. (2012). VizieR Online Data Catalog: 2MASS 6X Point Source Working Database / Catalog (Cutri+ 2006). *VizieR Online Data Catalog*, II/281. (in p. 29)

Dage, K.C., Zepf, S.E., Thygesen, E., Bahramian, A., Kundu, A., Maccarone, T.J., Peacock, M.B. & Strader, J. (2020). X-ray spectroscopy of newly identified ULXs associated with M87's globular cluster population. MNRAS, **497**, 596–608. (in p. 98)

Dale, D.A. & Helou, G. (2002). The Infrared Spectral Energy Distribution of Normal Star-forming Galaxies: Calibration at Far-Infrared and Submillimeter Wavelengths. ApJ, **576**, 159–168. (in pp. 35 and 69)

Dálya, G., Galgóczi, G., Dobos, L., Frei, Z., Heng, I.S., Macas, R., Messenger, C., Raffai, P. & de Souza, R.S. (2018). GLADE: A galaxy catalogue for multimessenger searches in the advanced gravitational-wave detector era. MNRAS, **479**, 2374–2381. (in pp. 16, 36, 37, 38, 43, and 45)

Das, A., Mesinger, A., Pallottini, A., Ferrara, A. & Wise, J.H. (2017). High-mass X-ray binaries and the cosmic 21-cm signal: impact of host galaxy absorption. MNRAS, **469**, 1166–1174. (in p. 64)

David, L.P., Jones, C., Forman, W. & Murray, S.S. (2005). The Diffuse Emission and a Variable Ultraluminous X-Ray Point Source in the Elliptical Galaxy NGC 3379. ApJ, **635**, 1053–1061. (in p. 98)

de Vaucouleurs, G. (1991). Book-Review - Third Reference Catalogue of Bright Galaxies VOL.1. *Science*, **254**, 592. (in pp. 17 and 70)

de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, J.R. (1976). Second reference catalogue of bright galaxies. In *Second reference catalogue of bright galaxies, Vol. 1976, p. Austin: University of Texas Press.*, vol. 1976. (in pp. 60 and 72)

Del Pozzo, W., Berry, C.P.L., Ghosh, A., Haines, T.S.F., Singer, L.P. & Vecchio, A. (2018). Dirichlet process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations. MNRAS, **479**, 601–614. (in p. 43)

Delvecchio, I., Daddi, E., Aird, J., Mullaney, J.R., Bernhard, E., Grimmett, L.P., Carraro, R., Cimatti, A., Zamorani, G., Caplar, N., Vito, F., Elbaz, D. & Rodighiero, G. (2020). The Evolving AGN Duty Cycle in Galaxies Since z ∼ 3 as Encoded in the X-Ray Luminosity Function. ApJ, **892**, 17. (in p. 80)

Deng, L.C., Newberg, H.J., Liu, C., Carlin, J.L., Beers, T.C., Chen, L., Chen, Y.Q., Christlieb, N., Grillmair, C.J., Guhathakurta, P., Han, Z.W., Hou, J.L., Lee, H.T., Lépine, S., Li, J., Liu, X.W., Pan, K.K., Sellwood, J.A., Wang, B., Wang, H.C., Yang, F., Yanny, B., Zhang, H.T., Zhang, Y.Y., Zheng, Z. & Zhu, Z. (2012). LAMOST Experiment for Galactic Understanding and Exploration (LEGUE) — The survey's science plan. *Research in Astronomy and Astrophysics*, **12**, 735–754. (in p. 42)

Douna, V.M., Pellizza, L.J., Mirabel, I.F. & Pedrosa, S.E. (2015). Metallicity dependence of high-mass X-ray binary populations. A&A, **579**, A44. (in pp. 8, 10, 11, and 65)

Ducoin, J.G., Corre, D., Leroy, N. & Le Floch, E. (2020). Optimizing gravitational waves follow-up using galaxies stellar mass. MNRAS, **492**, 4768–4779. (in pp. 16, 36, 43, 44, 45, and 47)

Earnshaw, H.P., Roberts, T.P., Middleton, M.J., Walton, D.J. & Mateos, S. (2019). A new, clean catalogue of extragalactic non-nuclear X-ray sources in nearby galaxies. MNRAS, **483**, 5554–5573. (in pp. 66, 70, 83, and 92)

Elmegreen, B.G. & Elmegreen, D.M. (2001). Fractal Structure in Galactic Star Fields. AJ, **121**, 1507–1511. (in p. 115)

Elmegreen, B.G. & Falgarone, E. (1996). A Fractal Origin for the Mass Spectrum of Interstellar Clouds. ApJ, **471**, 816. (in p. 73)

Elmegreen, D.M., Elmegreen, B.G., Marcus, M.T., Shahinyan, K., Yau, A. & Petersen, M. (2009). Clumpy Galaxies in Goods and Gems: Massive Analogs of Local Dwarf Irregulars. ApJ, **701**, 306–329. (in p. 115)

Eracleous, M., Shields, J.C., Chartas, G. & Moran, E.C. (2002). Three LINERs under the Chandra X-Ray Microscope. ApJ, **565**, 108–124. (in p. 80)

Evans, I.N., Primini, F.A., Glotfelty, K.J., Anderson, C.S., Bonaventura, N.R., Chen, J.C., Davis, J.E., Doe, S.M., Evans, J.D., Fabbiano, G., Galle, E.C., Gibbs, D.G., II, Grier, J.D., Hain, R.M., Hall, D.M., Harbo, P.N., (Helen He, X., Houck, J.C., Karovska, M., Kashyap, V.L., Lauer, J., McCollough, M.L., McDowell, J.C., Miller, J.B., Mitschang, A.W., Morgan, D.L., Mossman, A.E., Nichols, J.S., Nowak, M.A., Plummer, D.A., Refsdal, B.L., Rots, A.H., Siemiginowska, A., Sundheim, B.A., Tibbetts, M.S., Van Stone, D.W., Winkelman, S.L. & Zografou, P. (2010). The Chandra Source Catalog. ApJS, **189**, 37–82. (in pp. 42 and 70)

Fabbiano, G. (1989). X rays from normal galaxies. ARA&A, **27**, 87–138. (in pp. 9 and 63)

Fabbiano, G., Zezas, A. & Murray, S.S. (2001). Chandra Observations of "The Antennae" Galaxies (NGC 4038/9). ApJ, **554**, 1035–1043. (in pp. 11 and 65)

Fabbiano, G., Kim, D.W., Fragos, T., Kalogera, V., King, A.R., Angelini, L., Davies, R.L., Gallagher, J.S., Pellegrini, S., Trinchieri, G., Zepf, S.E. & Zezas, A. (2006). The Modulated Emission of the Ultraluminous X-Ray Source in NGC 3379. ApJ, **650**, 879–884. (in pp. 65 and 116)

Farrell, S.A., Webb, N.A., Barret, D., Godet, O. & Rodrigues, J.M. (2009). An intermediate-mass black hole of over 500 solar masses in the galaxy ESO243-49. Nature, **460**, 73–75. (in p. 63)

Feng, H. & Kaaret, P. (2008). Optical Counterpart of the Ultraluminous X-Ray Source IC 342 X-1. ApJ, **675**, 1067–1075. (in pp. 11, 64, and 65)

Finke, J.D. & Razzaque, S. (2017). The binary black hole merger rate from ultraluminous X-ray source progenitors. MNRAS, **472**, 3683–3691. (in pp. 10 and 64)

Flaugher, B. (2005). The Dark Energy Survey. *International Journal of Modern Physics A*, **20**, 3121–3123. (in p. 42)

Fornasini, F.M., Kriek, M., Sanders, R.L., Shivaei, I., Civano, F., Reddy, N.A., Shapley, A.E., Coil, A.L., Mobasher, B., Siana, B., Aird, J., Azadi, M., Freeman, W.R., Leung, G.C.K., Price, S.H., Fetherolf, T., Zick, T. & Barro, G. (2019). The MOSDEF Survey: The Metallicity Dependence of X-Ray Binary Populations at z 2. ApJ, **885**, 65. (in p. 97)

Fornasini, F.M., Civano, F. & Suh, H. (2020). Connecting the metallicity dependence and redshift evolution of high-mass X-ray binaries. MNRAS. (in pp. 10 and 97)

Fragos, T., Kalogera, V., Belczynski, K., Fabbiano, G., Kim, D.W., Brassington, N.J., Angelini, L., Davies, R.L., Gallagher, J.S., King, A.R., Pellegrini, S., Trinchieri, G., Zepf, S.E., Kundu, A. & Zezas, A. (2008). Models for Low-Mass X-Ray Binaries in the Elliptical Galaxies NGC 3379 and NGC 4278: Comparison with Observations. ApJ, **683**, 346–356. (in p. 8)

Fragos, T., Lehmer, B., Tremmel, M., Tzanavaris, P., Basu-Zych, A., Belczynski, K., Hornschemeier, A., Jenkins, L., Kalogera, V., Ptak, A. & Zezas, A. (2013a). X-Ray Binary Evolution Across Cosmic Time. ApJ, **764**, 41. (in pp. 8, 65, and 93)

Fragos, T., Lehmer, B.D., Naoz, S., Zezas, A. & Basu-Zych, A. (2013b). Energy Feedback from X-Ray Binaries in the Early Universe. ApJ, **776**, L31. (in pp. 8, 65, and 94)

Fragos, T., Linden, T., Kalogera, V. & Sklias, P. (2015). On the Formation of Ultraluminous X-Ray Sources with Neutron Star Accretors: The Case of M82 X-2. ApJ, **802**, L5. (in p. 64)

Freedman, W.L., Madore, B.F., Gibson, B.K., Ferrarese, L., Kelson, D.D., Sakai, S., Mould, J.R., Kennicutt, R.C., Jr., Ford, H.C., Graham, J.A., Huchra, J.P., Hughes, S.M.G., Illingworth, G.D., Macri, L.M. & Stetson, P.B. (2001). Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. ApJ, **553**, 47–72. (in p. 27)

French, K.D., Wevers, T., Law-Smith, J., Graur, O. & Zabludoff, A.I. (2020). The Host Galaxies of Tidal Disruption Events. Space Sci. Rev., **216**, 32. (in p. 49)

Fürst, F., Walton, D.J., Harrison, F.A., Stern, D., Barret, D., Brightman, M., Fabian, A.C., Grefenstette, B., Madsen, K.K., Middleton, M.J., Miller, J.M., Pottschmidt, K., Ptak, A., Rana, V. & Webb, N. (2016). Discovery of Coherent Pulsations from the Ultraluminous X-Ray Source NGC 7793 P13. ApJ, **831**, L14. (in pp. 9 and 64)

Galliano, F., Galametz, M. & Jones, A.P. (2018). The Interstellar Dust Properties of Nearby Galaxies. ARA&A, **56**, 673–713. (in p. 33)

Gao, Y., Wang, Q.D., Appleton, P.N. & Lucas, R.A. (2003). Nonnuclear Hyper/Ultraluminous X-Ray Sources in the Starbursting Cartwheel Ring Galaxy. ApJ, **596**, L171–L174. (in pp. 11 and 65)

Gardner, J.P., Sharples, R.M., Frenk, C.S. & Carrasco, B.E. (1997). A Wide-Field K-band Survey: The Luminosity Function of Galaxies. ApJ, **480**, L99–L102. (in p. 28)

Gehrels, N., Cannizzo, J.K., Kanner, J., Kasliwal, M.M., Nissanke, S. & Singer, L.P. (2016). Galaxy Strategy for LIGO-Virgo Gravitational Wave Counterpart Searches. ApJ, **820**, 136. (in pp. 16, 37, 38, 39, 43, and 45)

Ghosh, H., Mathur, S., Fiore, F. & Ferrarese, L. (2008). Low-Level Nuclear Activity in Nearby Spiral Galaxies. ApJ, **687**, 216–229. (in p. 80)

Bibliography

Giacconi, R., Gursky, H., Paolini, F.R. & Rossi, B.B. (1962). Evidence for x Rays From Sources Outside the Solar System. Phys. Rev. Lett., **9**, 439–443. (in p. 1)

Giacconi, R., Kellogg, E., Gorenstein, P., Gursky, H. & Tananbaum, H. (1971). An X-Ray Scan of the Galactic Plane from UHURU. ApJ, **165**, L27. (in p. 1)

Gilfanov, M. (2004). Low-mass X-ray binaries as a stellar mass indicator for the host galaxy. MNRAS, **349**, 146–168. (in pp. 5, 7, 8, and 116)

Gladstone, J.C., Roberts, T.P. & Done, C. (2009). The ultraluminous state. MNRAS, **397**, 1836–1851. (in pp. 9 and 64)

Goldstein, A., Veres, P., Burns, E., Briggs, M.S., Hamburg, R., Kocevski, D., Wilson-Hodge, C.A., Preece, R.D., Poolakkil, S., Roberts, O.J., Hui, C.M., Connaughton, V., Racusin, J., von Kienlin, A., Dal Canton, T., Christensen, N., Littenberg, T., Siellez, K., Blackburn, L., Broida, J., Bissaldi, E., Cleveland, W.H., Gibby, M.H., Giles, M.M., Kippen, R.M., McBreen, S., McEnery, J., Meegan, C.A., Paciesas, W.S. & Stanbro, M. (2017). An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A. ApJ, **848**, L14. (in p. 48)

Gong, H., Liu, J. & Maccarone, T. (2016). An Extreme Luminous X-ray Source Catalog Based on Chandra ACIS Observations. ApJS, **222**, 12. (in p. 80)

González Delgado, R.M., García-Benito, R., Pérez, E., Cid Fernandes, R., de Amorim, A.L., Cortijo-Ferrero, C., Lacerda, E.A.D., López Fernández, R., Vale-Asari, N., Sánchez, S.F., Mollá, M., Ruiz-Lara, T., Sánchez-Blázquez, P., Walcher, C.J., Alves, J., Aguerri, J.A.L., Bekeraité, S., Bland-Hawthorn, J., Galbany, L., Gallazzi, A., Husemann, B., Iglesias-Páramo, J., Kalinova, V., López-Sánchez, A.R., Marino, R.A., Márquez, I., Masegosa, J., Mast, D., Méndez-Abreu, J., Mendoza, A., del Olmo, A., Pérez, I., Quirrenbach, A. & Zibetti, S. (2015). The CALIFA survey across the Hubble sequence. Spatially resolved stellar population properties in galaxies. A&A, **581**, A103. (in p. 93)

Górski, K.M., Hivon, E., Banday, A.J., Wand elt, B.D., Hansen, F.K., Reinecke, M. & Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, **622**, 759–771. (in p. 44)

Greggio, L. & Cappellaro, E. (2019). Correlation of the rate of Type Ia supernovae with the parent galaxy properties: Light and shadows. A&A, **625**, A113. (in p. 50)

Greiner, J., Cuby, J.G., McCaughrean, M.J., Castro-Tirado, A.J. & Mennickent, R.E. (2001). Identification of the donor in the X-ray binary GRS 1915+105. A&A, **373**, L37–L40. (in p. 89)

Grimm, H.J., Gilfanov, M. & Sunyaev, R. (2003). High-mass X-ray binaries as a star formation rate indicator in distant galaxies. MNRAS, **339**, 793–809. (in pp. 5, 73, 111, and 116)

Hao, C.N., Kennicutt, R.C., Johnson, B.D., Calzetti, D., Dale, D.A. & Moustakas, J. (2011). Dust-corrected Star Formation Rates of Galaxies. II. Combinations of Ultraviolet and Infrared Tracers. ApJ, **741**, 124. (in p. 28)

Hawkins, E., Maddox, S., Cole, S., Lahav, O., Madgwick, D.S., Norberg, P., Peacock, J.A., Baldry, I.K., Baugh, C.M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Dalton, G., De Propris, R., Driver, S.P., Efstathiou, G., Ellis, R.S., Frenk, C.S., Glazebrook, K., Jackson, C., Jones, B., Lewis, I., Lumsden, S., Percival, W., Peterson, B.A., Sutherland, W. & Taylor, K. (2003). The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. MNRAS, **346**, 78–96. (in p. 25)

Hayward, C.C., Lanz, L., Ashby, M.L.N., Fazio, G., Hernquist, L., Martínez-Galarza, J.R., Noeske, K., Smith, H.A., Wuyts, S. & Zezas, A. (2014). The total infrared luminosity may significantly overestimate the star formation rate of quenching and recently quenched galaxies. MNRAS, **445**, 1598–1604. (in p. 41)

He, H.N., Kusenko, A., Nagataki, S., Zhang, B.B., Yang, R.Z. & Fan, Y.Z. (2016). Monte Carlo Bayesian search for the plausible source of the Telescope Array hotspot. Phys. Rev. D, **93**, 043011. (in p. 49)

Heger, A., Fryer, C.L., Woosley, S.E., Langer, N. & Hartmann, D.H. (2003). How Massive Single Stars End Their Life. ApJ, **591**, 288–300. (in pp. 12, 65, and 95)

Helou, G. & Walker, D.W. (1988). Infrared Astronomical Satellite (IRAS), Catalogs and Atlases. Vol. 7. The Small Scale Structure Catalog. In *NASA RP-1190, Vol. 7 (1988)*, vol. 7. (in pp. 35 and 69)

Helou, G., Madore, B.F., Schmitz, M., Bicay, M.D., Wu, X. & Bennett, J. (1991). The NASA/IPAC extragalactic database. In M.A. Albrecht & D. Egret, eds., *Databases and On-line Data in Astronomy*, vol. 171 of *Astrophysics and Space Science Library*, 89–106. (in p. 15)

Ho, L.C., Feigelson, E.D., Townsley, L.K., Sambruna, R.M., Garmire, G.P., Brandt, W.N., Filippenko, A.V., Griffiths, R.E., Ptak, A.F. & Sargent, W.L.W. (2001). Detection of Nuclear X-Ray Sources in Nearby Galaxies with Chandra. ApJ, **549**, L51–L54. (in pp. 4 and 80)

IceCube Collaboration (2018). Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A. *Science*, **361**, eaat1378. (in p. 49)

Irwin, J.A., Bregman, J.N. & Athey, A.E. (2004). The Lack of Very Ultraluminous X-Ray Sources in Early-Type Galaxies. ApJ, **601**, L143–L146. (in p. 11)

Israel, G.L., Belfiore, A., Stella, L., Esposito, P., Casella, P., De Luca, A., Marelli, M., Papitto, A., Perri, M., Puccetti, S., Castillo, G.A.R., Salvetti, D., Tiengo, A., Zampieri, L., D'Agostino, D., Greiner, J., Haberl, F., Novara, G., Salvaterra, R., Turolla, R., Watson, M., Wilms, J. & Wolter, A. (2017a). An accreting pulsar with extreme properties drives an ultraluminous x-ray source in NGC 5907. *Science*, **355**, 817–819. (in pp. 9 and 64)

Israel, G.L., Papitto, A., Esposito, P., Stella, L., Zampieri, L., Belfiore, A., Rodríguez Castillo, G.A., De Luca, A., Tiengo, A., Haberl, F., Greiner, J., Salvaterra, R., Sandrelli, S. & Lisini, G. (2017b). Discovery of a 0.42-s pulsar in the ultraluminous X-ray source NGC 7793 P13. MNRAS, **466**, L48–L52. (in pp. 9 and 64)

Bibliography

Ivezić, Ž., Kahn, S.M., Tyson, J.A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S.F., Andrew, J., Angel, J.R.P., Angeli, G.Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K.T., Astier, P., Aubourg, É., Auza, N., Axelrod, T.S., Bard, D.J., Barr, J.D., Barrau, A., Bartlett, J.G., Bauer, A.E., Bauman, B.J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A.C., Becla, J., Beldica, C., Bellavia, S., Bianco, F.B., Biswas, R., Blanc, G., Blazek, J., Bland ford, R.D., Bloom, J.S., Bogart, J., Bond, T.W., Booth, M.T., Borgland, A.W., Borne, K., Bosch, J.F., Boutigny, D., Brackett, C.A., Bradshaw, A., Brand t, W.N., Brown, M.E., Bullock, J.S., Burchat, P., Burke, D.L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A.L., Carlin, J.L., Carlson, E.L., Chand rasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E.C., Chiang, H.F., Chiang, J., Chirino, C., Chow, D., Ciardi, D.R., Claver, C.F., Cohen-Tanugi, J., Cockrum, J.J., Coles, R., Connolly, A.J., Cook, K.H., Cooray, A., Covey, K.R., Cribbs, C., Cui, W., Cutri, R., Daly, P.N., Daniel, S.F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S.W., Doherty, P., Dubois, R., Dubois-Felsmann, G.P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B.L., Fausti Neto, A., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M.D., Frank, J., Freemon, M.D., Gangler, E., Gawiser, E., Geary, J.C., Gee, P., Geha, M., Gessner, C.J.B., Gibson, R.R., Gilmore, D.K., Glanzman, T., Glick, W., Goldina, T., Goldstein, D.A., Goodenow, I., Graham, M.L., Gressler, W.J., Gris, P., Guy, L.P., Guyonnet, A., Haller, G., Harris, R., Hascall, P.A., Haupt, J., Hernand ez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J.A., Hogan, C., Howard, J.D., Huang, D., Huffer, M.E., Ingraham, P., Innes, W.R., Jacoby, S.H., Jain, B., Jammes, F., Jee, M.J., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A.S., Johnson, M.W.G., Jones, R.L., Juramy-Gilles, C., Jurić, M., Kalirai, J.S., Kallivayalil, N.J., Kalmbach, B., Kantor, J.P., Karst, P., Kasliwal, M.M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I.V., Krabbendam, V.L., Krughoff, K.S., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N.R., Lage, C.S., Lambert, R., Lange, T., Langton, J.B., Le Guillou, L., Levine, D., Liang, M., Lim, K.T., Lintott, C.J., Long, K.E., Lopez, M., Lotz, P.J., Lupton, R.H., Lust, N.B., MacArthur, L.A., Mahabal, A., Mand elbaum, R., Markiewicz, T.W., Marsh, D.S., Marshall, P.J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., Mills, D.J., Miraval, C., Moeyens, J., Moolekamp, F.E., Monet, D.G., Moniez, M., Monkewitz, S., Montgomery, C., Morrison, C.B., Mueller, F., Muller, G.P., Muñoz Arancibia, F., Neill, D.R., Newbry, S.P., Nief, J.Y., Nomerotski, A., Nordby, M., O'Connor, P., Oliver, J., Olivier, S.S., Olsen, K., O'Mullane, W., Ortiz, S., Osier, S., Owen, R.E., Pain, R., Palecek, P.E., Parejko, J.K., Parsons, J.B., Pease, N.M., Peterson, J.M., Peterson, J.R., Petravick, D.L., Libby Petrick, M.E., Petry, C.E., Pierfederici, F., Pietrowicz, S., Pike, R., Pinto, P.A., Plante, R., Plate, S., Plutchak, J.P., Price, P.A., Prouza, M., Radeka, V., Rajagopal, J., Rasmussen, A.P., Regnault, N., Reil, K.A., Reiss, D.J., Reuter, M.A., Ridgway, S.T., Riot, V.J., Ritz, S., Robinson, S., Roby, W., Roodman, A., Rosing, W., Roucelle, C., Rumore, M.R., Russo, S., Saha, A., Sassolas, B., Schalk, T.L., Schellart, P., Schindler, R.H., Schmidt, S., Schneider, D.P., Schneider, M.D., Schoening, W., Schumacher, G., Schwamb, M.E.,

Sebag, J., Selvy, B., Sembroski, G.H., Seppala, L.G., Serio, A., Serrano, E., Shaw, R.A., Shipsey, I., Sick, J., Silvestri, N., Slater, C.T., Smith, J.A., Smith, R.C., Sobhani, S., Soldahl, C., Storrie-Lombardi, L., Stover, E., Strauss, M.A., Street, R.A., Stubbs, C.W., Sullivan, I.S., Sweeney, D., Swinbank, J.D., Szalay, A., Takacs, P., Tether, S.A., Thaler, J.J., Thayer, J.G., Thomas, S., Thornton, A.J., Thukral, V., Tice, J., Trilling, D.E., Turri, M., Van Berg, R., Vanden Berk, D., Vetter, K., Virieux, F., Vucina, T., Wahl, W., Walkowicz, L., Walsh, B., Walter, C.W., Wang, D.L., Wang, S.Y., Warner, M., Wiecha, O., Willman, B., Winters, S.E., Wittman, D., Wolff, S.C., Wood-Vasey, W.M., Wu, X., Xin, B., Yoachim, P. & Zhan, H. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products. ApJ, **873**, 111. (in p. 42)

Jarrett, T.H., Chester, T., Cutri, R., Schneider, S., Skrutskie, M. & Huchra, J.P. (2000). 2MASS Extended Source Catalog: Overview and Algorithms. AJ, **119**, 2498–2531. (in p. 56)

Jarrett, T.H., Chester, T., Cutri, R., Schneider, S.E. & Huchra, J.P. (2003). The 2MASS Large Galaxy Atlas. AJ, **125**, 525–554. (in p. 29)

Jones, D.H., Saunders, W., Colless, M., Read, M.A., Parker, Q.A., Watson, F.G., Campbell, L.A., Burkey, D., Mauch, T., Moore, L., Hartley, M., Cass, P., James, D., Russell, K., Fiegert, K., Dawe, J., Huchra, J., Jarrett, T., Lahav, O., Lucey, J., Mamon, G.A., Proust, D., Sadler, E.M. & Wakamatsu, K.i. (2004). The 6dF Galaxy Survey: samples, observational techniques and the first data release. MNRAS, **355**, 747–763. (in p. 56)

Kaaret, P. (2014). X-ray luminous binaries, metallicity, and the early Universe. MNRAS, **440**, L26–L30. (in p. 10)

Kaaret, P., Alonso-Herrero, A., Gallagher, J.S., Fabbiano, G., Zezas, A. & Rieke, M.J. (2004). Displacement of X-ray sources from star clusters in starburst galaxies. MNRAS, **348**, L28–L32. (in p. 65)

Kaaret, P., Feng, H. & Roberts, T.P. (2017). Ultraluminous X-Ray Sources. ARA&A, **55**, 303–341. (in pp. 63 and 119)

Kalogera, V., Belczynski, K., Kim, C., O'Shaughnessy, R. & Willems, B. (2007). Formation of double compact objects. Phys. Rep., **442**, 75–108. (in p. 43)

Kanner, J., Huard, T.L., Márka, S., Murphy, D.C., Piscionere, J., Reed, M. & Shawhan, P. (2008). LOOC UP: locating and observing optical counterparts to gravitational wave bursts. *Classical and Quantum Gravity*, **25**, 184034. (in p. 43)

Karachentsev, I.D. & Makarov, D.A. (1996). The Galaxy Motion Relative to Nearby Galaxies and the Local Velocity Field. AJ, **111**, 794. (in p. 53)

Karachentsev, I.D., Karachentseva, V.E. & Huchtmeier, W.K. (2001). H I observations of nearby galaxies. III. More dwarf galaxies in the northern sky. A&A, **366**, 428–438. (in p. 56)

Karachentsev, I.D., Karachentseva, V.E. & Huchtmeier, W.K. (2007). New probable dwarf galaxies in northern groups of the Local Supercluster. *Astronomy Letters*, **33**, 512–519. (in p. 56)

Karachentsev, I.D., Makarov, D.I. & Kaisina, E.I. (2013). Updated Nearby Galaxy Catalog. AJ, **145**, 101. (in p. 56)

Kasliwal, M.M., Nakar, E., Singer, L.P., Kaplan, D.L., Cook, D.O., Van Sistine, A., Lau, R.M., Fremling, C., Gottlieb, O., Jencson, J.E., Adams, S.M., Feindt, U., Hotokezaka, K., Ghosh, S., Perley, D.A., Yu, P.C., Piran, T., Allison, J.R., Anupama, G.C., Balasubramanian, A., Bannister, K.W., Bally, J., Barnes, J., Barway, S., Bellm, E., Bhalerao, V., Bhattacharya, D., Blagorodnova, N., Bloom, J.S., Brady, P.R., Cannella, C., Chatterjee, D., Cenko, S.B., Cobb, B.E., Copperwheat, C., Corsi, A., De, K., Dobie, D., Emery, S.W.K., Evans, P.A., Fox, O.D., Frail, D.A., Frohmaier, C., Goobar, A., Hallinan, G., Harrison, F., Helou, G., Hinderer, T., Ho, A.Y.Q., Horesh, A., Ip, W.H., Itoh, R., Kasen, D., Kim, H., Kuin, N.P.M., Kupfer, T., Lynch, C., Madsen, K., Mazzali, P.A., Miller, A.A., Mooley, K., Murphy, T., Ngeow, C.C., Nichols, D., Nissanke, S., Nugent, P., Ofek, E.O., Qi, H., Quimby, R.M., Rosswog, S., Rusu, F., Sadler, E.M., Schmidt, P., Sollerman, J., Steele, I., Williamson, A.R., Xu, Y., Yan, L., Yatsu, Y., Zhang, C. & Zhao, W. (2017). Illuminating gravitational waves: A concordant picture of photons from a neutron star merger. *Science*, **358**, 1559–1565. (in pp. 43 and 45)

Kauffmann, G., Heckman, T.M., Tremonti, C., Brinchmann, J., Charlot, S., White, S.D.M., Ridgway, S.E., Brinkmann, J., Fukugita, M., Hall, P.B., Ivezić, Ž., Richards, G.T. & Schneider, D.P. (2003). The host galaxies of active galactic nuclei. MNRAS, **346**, 1055–1077. (in pp. 15, 30, and 69)

Kelly, B.C. (2007). Some Aspects of Measurement Error in Linear Regression of Astronomical Data. ApJ, **665**, 1489–1506. (in p. 25)

Kennicutt, R.C. & Evans, N.J. (2012). Star Formation in the Milky Way and Nearby Galaxies. ARA&A, **50**, 531–608. (in pp. 28, 31, 33, 35, 69, and 100)

Kennicutt, R.C., Jr. (1998). Star Formation in Galaxies Along the Hubble Sequence. ARA&A, **36**, 189–232. (in pp. 35 and 69)

Kewley, L.J. & Ellison, S.L. (2008). Metallicity Calibrations and the Mass-Metallicity Relation for Star-forming Galaxies. ApJ, **681**, 1183–1204. (in pp. 35, 95, and 97)

Kim, D.W. & Fabbiano, G. (2004). X-Ray Luminosity Function and Total Luminosity of Low-Mass X-Ray Binaries in Early-Type Galaxies. ApJ, **611**, 846–857. (in p. 65)

Kim, D.W. & Fabbiano, G. (2010). X-ray Properties of Young Early-type Galaxies. I. X-ray Luminosity Function of Low-mass X-ray Binaries. ApJ, **721**, 1523–1530. (in p. 98)

Kim, E., Kim, D.W., Fabbiano, G., Lee, M.G., Park, H.S., Geisler, D. & Dirsch, B. (2006). Low-Mass X-Ray Binaries in Six Elliptical Galaxies: Connection to Globular Clusters. ApJ, **647**, 276–292. (in p. 98)

Kim, M., Wilkes, B.J., Kim, D.W., Green, P.J., Barkhouse, W.A., Lee, M.G., Silverman, J.D. & Tananbaum, H.D. (2007). Chandra Multiwavelength Project X-Ray Point Source Number Counts and the Cosmic X-Ray Background. ApJ, **659**, 29–51. (in pp. 16 and 77)

Kim, S., Rey, S.C., Jerjen, H., Lisker, T., Sung, E.C., Lee, Y., Chung, J., Pak, M., Yi, W. & Lee, W. (2014). The Extended Virgo Cluster Catalog. ApJS, **215**, 22. (in p. 23)

King, A. & Lasota, J.P. (2016). ULXs: Neutron stars versus black holes. MNRAS, **458**, L10–L13. (in pp. 9 and 64)

King, A., Lasota, J.P. & Kluźniak, W. (2017). Pulsing ULXs: tip of the iceberg? MNRAS, **468**, L59–L62. (in pp. 9 and 64)

King, A.R. (2004). Ultraluminous X-ray sources and star formation. MNRAS, **347**, L18–L20. (in p. 9)

King, A.R. (2009). Masses, beaming and Eddington ratios in ultraluminous X-ray sources. MNRAS, **393**, L41–L44. (in p. 9)

King, A.R., Davies, M.B., Ward, M.J., Fabbiano, G. & Elvis, M. (2001). Ultraluminous X-Ray Sources in External Galaxies. ApJ, **552**, L109–L112. (in pp. 9 and 64)

Koliopanos, F., Vasilopoulos, G., Buchner, J., Maitra, C. & Haberl, F. (2019). Investigating ULX accretion flows and cyclotron resonance in NGC 300 ULX1. A&A, **621**, A118. (in p. 64)

Komis, I., Pavlidou, V. & Zezas, A. (2019). Extragalactic gamma-ray background from star-forming galaxies: Will empirical scalings suffice? MNRAS, **483**, 4020–4030. (in p. 16)

Kopparapu, R.K., Hanna, C., Kalogera, V., O'Shaughnessy, R., González, G., Brady, P.R. & Fairhurst, S. (2008). Host Galaxies Catalog Used in LIGO Searches for Compact Binary Coalescence Events. ApJ, **675**, 1459–1467. (in pp. 16, 38, 39, and 43)

Kourkchi, E. & Tully, R.B. (2017). Galaxy Groups Within 3500 km s^{-1}. ApJ, **843**, 16. (in p. 48)

Kouroumpatzakis, K., Zezas, A., Sell, P., Kovlakas, K., Bonfini, P., Willner, S.P., Ashby, M.L.N., Maragkoudakis, A. & Jarrett, T.H. (2020). Sub-galactic scaling relations between X-ray luminosity, star formation rate, and stellar mass. MNRAS, **494**, 5967–5984. (in pp. 69, 97, and 102)

Kovlakas, K., Zezas, A., Andrews, J.J., Basu-Zych, A., Fragos, T., Hornschemeier, A., Lehmer, B. & Ptak, A. (2020). A census of ultraluminous X-ray sources in the local Universe. MNRAS, **498**, 4790–4810. (in pp. 16, 34, 42, 119, 136, and 138)

Krauss, F., Calamari, E., Keivani, A., Coleiro, A., Evans, P.A., Fox, D.B., Kennea, J.A., Mészáros, P., Murase, K., Russell, T.D., Santander, M. & Tohuvavohu, A. (2020). Multimessenger observations of counterparts to IceCube-190331A. MNRAS, **497**, 2553–2561. (in p. 49)

Kuranov, A.G., Popov, S.B., Postnov, K.A., Volonteri, M. & Perna, R. (2007). On the dynamical formation of accreting intermediate mass black holes. MNRAS, **377**, 835–842. (in p. 63)

Lang, D. (2014). unWISE: Unblurred Coadds of the WISE Imaging. AJ, **147**, 108. (in p. 29)

Lang, D., Hogg, D.W. & Schlegel, D.J. (2016). WISE Photometry for 400 Million SDSS Sources. AJ, **151**, 36. (in pp. 29, 32, 41, and 69)

Lehmer, B.D., Brandt, W.N., Alexander, D.M., Bauer, F.E., Conselice, C.J., Dickinson, M.E., Giavalisco, M., Grogin, N.A., Koekemoer, A.M., Lee, K.S., Moustakas, L.A. & Schneider, D.P. (2005). X-Ray Properties of Lyman Break Galaxies in the Great Observatories Origins Deep Survey. AJ, **129**, 1–8. (in pp. 10 and 97)

Lehmer, B.D., Basu-Zych, A.R., Mineo, S., Brandt, W.N., Eufrasio, R.T., Fragos, T., Hornschemeier, A.E., Luo, B., Xue, Y.Q., Bauer, F.E., Gilfanov, M., Ranalli, P., Schneider, D.P., Shemmer, O., Tozzi, P., Trump, J.R., Vignali, C., Wang, J.X., Yukita, M. & Zezas, A. (2016). The Evolution of Normal Galaxy X-Ray Emission through Cosmic History: Constraints from the 6 MS Chandra Deep Field-South. ApJ, **825**, 7. (in pp. 65 and 97)

Lehmer, B.D., Eufrasio, R.T., Markwardt, L., Zezas, A., Basu-Zych, A., Fragos, T., Hornschemeier, A.E., Ptak, A., Tzanavaris, P. & Yukita, M. (2017). On the Spatially Resolved Star Formation History in M51. II. X-Ray Binary Population Evolution. ApJ, **851**, 11. (in p. 8)

Lehmer, B.D., Eufrasio, R.T., Tzanavaris, P., Basu-Zych, A., Fragos, T., Prestwich, A., Yukita, M., Zezas, A., Hornschemeier, A.E. & Ptak, A. (2019). X-Ray Binary Luminosity Function Scaling Relations for Local Galaxies Based on Subgalactic Modeling. ApJS, **243**, 3. (in pp. 7, 8, 94, 97, 101, 116, and 137)

Lehmer, B.D., Eufrasio, R.T., Basu-Zych, A., Doore, K., Fragos, T., Garofali, K., Kovlakas, K., Williams, B.J., Zezas, A. & Santana-Silva, L. (2020). The Metallicity Dependence of the High-Mass X-ray Binary Luminosity Function. *arXiv e-prints*, arXiv:2011.09476. (in p. 8)

Leibler, C.N. & Berger, E. (2010). The Stellar Ages and Masses of Short Gamma-ray Burst Host Galaxies: Investigating the Progenitor Delay Time Distribution and the Role of Mass and Star Formation in the Short Gamma-ray Burst Rate. ApJ, **725**, 1202–1214. (in p. 48)

Linden, T., Kalogera, V., Sepinsky, J.F., Prestwich, A., Zezas, A. & Gallagher, J.S. (2010). The Effect of Starburst Metallicity on Bright X-ray Binary Formation Pathways. ApJ, **725**, 1984–1994. (in pp. 8, 12, 65, and 95)

Liu, J. (2011). Chandra ACIS Survey of X-ray Point Sources in 383 Nearby Galaxies. I. The Source Catalog. ApJS, **192**, 10. (in p. 112)

Liu, J.F., Bregman, J.N. & Irwin, J. (2006). Ultraluminous X-Ray Sources in Nearby Galaxies from ROSAT HRI Observations. II. Statistical Properties. ApJ, **642**, 171–187. (in pp. 65 and 83)

Long, K.S. & van Speybroeck, L.P. (1983). X-ray emission from normal galaxies. In W.H.G. Lewin & E.P.J. van den Heuvel, eds., *Accretion-Driven Stellar X-ray Sources*, 117–146. (in pp. 9 and 63)

Luangtip, W., Roberts, T.P., Mineo, S., Lehmer, B.D., Alexander, D.M., Jackson, F.E., Goulding, A.D. & Fischer, J.L. (2015). A deficit of ultraluminous X-ray sources in luminous infrared galaxies. MNRAS, **446**, 470–492. (in p. 11)

Maccarone, T.J., Kundu, A., Zepf, S.E. & Rhode, K.L. (2007). A black hole in a globular cluster. Nature, **445**, 183–185. (in p. 98)

Madau, P. & Dickinson, M. (2014). Cosmic Star-Formation History. ARA&A, **52**, 415–486. (in pp. 40 and 97)

Madau, P. & Fragos, T. (2017). Radiation Backgrounds at Cosmic Dawn: X-Rays from Compact Binaries. ApJ, **840**, 39. (in pp. 10 and 64)

Madau, P., Rees, M.J., Volonteri, M., Haardt, F. & Oh, S.P. (2004). Early Reionization by Miniquasars. ApJ, **604**, 484–494. (in pp. 10 and 64)

Makarov, D., Prugniel, P., Terekhova, N., Courtois, H. & Vauglin, I. (2014). HyperLEDA. III. The catalogue of extragalactic distances. A&A, **570**, A13. (in pp. 13, 15, 17, 56, and 67)

Makishima, K., Kubota, A., Mizuno, T., Ohnishi, T., Tashiro, M., Aruga, Y., Asai, K., Dotani, T., Mitsuda, K., Ueda, Y., Uno, S., Yamaoka, K., Ebisawa, K., Kohmura, Y. & Okada, K. (2000). The Nature of Ultraluminous Compact X-Ray Sources in Nearby Spiral Galaxies. ApJ, **535**, 632–643. (in p. 63)

Mapelli, M., Ripamonti, E., Zampieri, L., Colpi, M. & Bressan, A. (2010). Ultra-luminous X-ray sources and remnants of massive metal-poor stars. MNRAS, **408**, 234–253. (in pp. 8, 11, 65, 95, 96, 97, and 117)

Mapelli, M., Ripamonti, E., Zampieri, L. & Colpi, M. (2011). Remnants of massive metal-poor stars: Viable engines for ultra-luminous X-ray sources. *Astronomische Nachrichten*, **332**, 414. (in p. 65)

Mapelli, M., Giacobbo, N., Ripamonti, E. & Spera, M. (2017). The cosmic merger rate of stellar black hole binaries from the Illustris simulation. MNRAS, **472**, 2422–2435. (in p. 43)

Mapelli, M., Giacobbo, N., Toffano, M., Ripamonti, E., Bressan, A., Spera, M. & Branchesi, M. (2018). The host galaxies of double compact objects merging in the local Universe. MNRAS, **481**, 5324–5330. (in p. 43)

Maragkoudakis, A., Zezas, A., Ashby, M.L.N. & Willner, S.P. (2017). The sub-galactic and nuclear main sequences for local star-forming galaxies. MNRAS, **466**, 1192–1204. (in p. 90)

Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T.M., de Mink, S., Mandel, I. & Moriya, T.J. (2017). Ultra-luminous X-ray sources and neutron-star-black-hole mergers from very massive close binaries at low metallicity. A&A, **604**, A55. (in pp. 10, 64, 65, and 95)

Marcote, B., Nimmo, K., Hessels, J.W.T., Tendulkar, S.P., Bassa, C.G., Paragi, Z., Keimpema, A., Bhardwaj, M., Karuppusamy, R., Kaspi, V.M., Law, C.J., Michilli, D., Aggarwal, K., Andersen, B., Archibald, A.M., Bandura, K., Bower, G.C., Boyle, P.J., Brar, C., Burke-Spolaor, S., Butler, B.J., Cassanelli, T., Chawla, P., Demorest, P., Dobbs, M., Fonseca, E., Giri, U., Good, D.C., Gourdji, K., Josephy, A., Kirichenko, A.Y., Kirsten, F., Landecker, T.L., Lang, D., Lazio, T.J.W., Li, D.Z., Lin, H.H., Linford, J.D., Masui, K., Mena-Parra, J., Naidu, A., Ng, C., Patel, C., Pen, U.L., Pleunis, Z., Rafiei-Ravandi, M., Rahman, M., Renard, A., Scholz, P., Siegel, S.R., Smith, K.M., Stairs, I.H., Vanderlinde, K. & Zwaniga, A.V. (2020). A repeating fast radio burst source localized to a nearby spiral galaxy. Nature, **577**, 190–194. (in p. 50)

McDermid, R.M., Alatalo, K., Blitz, L., Bournaud, F., Bureau, M., Cappellari, M., Crocker, A.F., Davies, R.L., Davis, T.A., de Zeeuw, P.T., Duc, P.A., Emsellem, E., Khochfar, S., Krajnović, D., Kuntschner, H., Morganti, R., Naab, T., Oosterloo, T., Sarzi, M., Scott, N., Serra, P., Weijmans, A.M. & Young, L.M. (2015). The ATLAS^{3D} Project - XXX. Star formation histories and stellar population scaling relations of early-type galaxies. MNRAS, **448**, 3484–3513. (in pp. 99, 100, 101, and 137)

Merloni, A., Predehl, P., Becker, W., Böhringer, H., Boller, T., Brunner, H., Brusa, M., Dennerl, K., Freyberg, M., Friedrich, P., Georgakakis, A., Haberl, F., Hasinger, G., Meidinger, N., Mohr, J., Nandra, K., Rau, A., Reiprich, T.H., Robrade, J., Salvato, M., Santangelo, A., Sasaki, M., Schwope, A., Wilms, J. & German eROSITA Consortium, t. (2012). eROSITA Science Book: Mapping the Structure of the Energetic Universe. *arXiv e-prints*, arXiv:1209.3114. (in p. 104)

Middleton, M.J. & King, A. (2017). Predicting ultraluminous X-ray source demographics from geometrical beaming. MNRAS, **470**, L69–L71. (in pp. 9 and 64)

Middleton, M.J., Heil, L., Pintore, F., Walton, D.J. & Roberts, T.P. (2015). A spectral-timing model for ULXs in the supercritical regime. MNRAS, **447**, 3243–3263. (in p. 64)

Mineo, S., Gilfanov, M. & Sunyaev, R. (2012). X-ray emission from star-forming galaxies - I. High-mass X-ray binaries. MNRAS, **419**, 2095–2115. (in pp. 5, 6, 8, 11, 73, 92, 111, 116, 117, 128, and 137)

Misra, D., Fragos, T., Tauris, T., Zapartas, E. & Aguilera-Dena, D.R. (2020). The Origin of Pulsating Ultra-Luminous X-ray Sources: Intermediate-Mass X-ray Binaries containing Neutron Star Accretors. *arXiv e-prints*, arXiv:2004.01205. (in pp. 9 and 64)

Mondal, S., Belczyński, K., Wiktorowicz, G., Lasota, J.P. & King, A.R. (2020). The connection between merging double compact objects and the ultraluminous X-ray sources. MNRAS, **491**, 2747–2759. (in p. 64)

Moretti, A., Poggianti, B.M., Fasano, G., Bettoni, D., D'Onofrio, M., Fritz, J., Cava, A., Varela, J., Vulcani, B., Gullieuszik, M., Couch, W.J., Omizzolo, A., Valentinuzzi, T., Dressler, A., Moles, M., Kjærgaard, P., Smareglia, R. & Molinaro, M. (2014). WINGS Data Release: a database of galaxies in nearby clusters. A&A, **564**, A138. (in p. 56)

Moshir, M.e. (1990). IRAS Faint Source Catalogue, version 2.0. In *IRAS Faint Source Catalogue, version 2.0 (1990)*. (in p. 29)

Mullaney, J.R., Alexander, D.M., Goulding, A.D. & Hickox, R.C. (2011). Defining the intrinsic AGN infrared spectral energy distribution and measuring its contribution to the infrared output of composite galaxies. MNRAS, **414**, 1082–1110. (in p. 80)

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability & Its Applications*, **9**, 141–142. (in p. 23)

Neijssel, C.J., Vigna-Gómez, A., Stevenson, S., Barrett, J.W., Gaebel, S.M., Broekgaarden, F.S., de Mink, S.E., Szécsi, D., Vinciguerra, S. & Mandel, I. (2019). The effect of the metallicity-specific star formation history on double compact object mergers. MNRAS, **490**, 3740–3759. (in p. 43)

Nilson, P. (1973). Uppsala general catalogue of galaxies. *Uppsala Astron. Obs. Ann.*, **6**. (in p. 17)

Nissanke, S., Kasliwal, M. & Georgieva, A. (2013). Identifying Elusive Electromagnetic Counterparts to Gravitational Wave Mergers: An End-to-end Simulation. ApJ, **767**, 124. (in p. 43)

Nuttall, L.K. & Sutton, P.J. (2010). Identifying the host galaxy of gravitational wave signals. Phys. Rev. D, **82**, 102002. (in p. 43)

O'Shaughnessy, R., Kim, C., Kalogera, V. & Belczynski, K. (2008). Constraining Population Synthesis Models via Empirical Binary Compact Object Merger and Supernova Rates. ApJ, **672**, 479–488. (in p. 4)

O'Shaughnessy, R., Bellovary, J.M., Brooks, A., Shen, S., Governato, F. & Christensen, C.R. (2017). The effects of host galaxy properties on merging compact binaries detectable by LIGO. MNRAS, **464**, 2831–2839. (in p. 43)

Pacucci, F., Mesinger, A., Mineo, S. & Ferrara, A. (2014). The X-ray spectra of the first galaxies: 21 cm signatures. MNRAS, **443**, 678–686. (in p. 10)

Park, T., Kashyap, V.L., Siemiginowska, A., van Dyk, D.A., Zezas, A., Heinke, C. & Wargelin, B.J. (2006). Bayesian Estimation of Hardness Ratios: Modeling and Computations. ApJ, **652**, 610–628. (in p. 77)

Parkash, V., Brown, M.J.I., Jarrett, T.H. & Bonne, N.J. (2018). Relationships between HI Gas Mass, Stellar Mass, and the Star Formation Rate of HICAT+WISE (H I-WISE) Galaxies. ApJ, **864**, 40. (in p. 33)

Pettini, M. & Pagel, B.E.J. (2004). [OIII]/[NII] as an abundance indicator at high redshift. MNRAS, **348**, L59–L63. (in pp. 35, 70, and 97)

Phinney, E.S. (1991). The Rate of Neutron Star Binary Mergers in the Universe: Minimal Predictions for Gravity Wave Detectors. ApJ, **380**, L17. (in p. 43)

Pierre Auger Collaboration (2010). Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astroparticle Physics*, **34**, 314–326. (in p. 49)

Pierre Auger Collaboration (2015). Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory. ApJ, **804**, 15. (in p. 49)

Pierre Auger Collaboration (2017). Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV. *Science*, **357**, 1266–1270. (in p. 49)

Pietrzyński, G., Graczyk, D., Gieren, W., Thompson, I.B., Pilecki, B., Udalski, A., Soszyński, I., Kozłowski, S., Konorski, P., Suchomska, K., Bono, G., Moroni, P.G.P., Villanova, S., Nardetto, N., Bresolin, F., Kudritzki, R.P., Storm, J., Gallenne, A., Smolec, R., Minniti, D., Kubiak, M., Szymański, M.K., Poleski, R., Wyrzykowski, Ł., Ulaczyk, K., Pietrukowicz, P., Górski, M. & Karczmarek, P. (2013). An eclipsing-binary distance to the Large Magellanic Cloud accurate to two per cent. Nature, **495**, 76–79. (in p. 19)

Pilyugin, L.S. (2001). On the oxygen abundance determination in HII regions. High-metallicity regions. A&A, **369**, 594–604. (in p. 97)

Pilyugin, L.S. & Thuan, T.X. (2005). Oxygen Abundance Determination in H II Regions: The Strong Line Intensities-Abundance Calibration Revisited. ApJ, **631**, 231–243. (in p. 97)

Planck Collaboration, Ade, P.A.R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A.J., Barreiro, R.B., Bartlett, J.G. & et al. (2016). Planck 2015 results. XIII. Cosmological parameters. A&A, **594**, A13. (in pp. 22 and 27)

Plotkin, R.M., Gallo, E., Miller, B.P., Baldassare, V.F., Treu, T. & Woo, J.H. (2014). An Environmental Study of the Ultraluminous X-Ray Source Population in Early-type Galaxies. ApJ, **780**, 6. (in pp. 11, 65, 98, 99, and 100)

Podsiadlowski, P. & Rappaport, S. (2000). Cygnus X-2: The Descendant of an Intermediate-Mass X-Ray Binary. ApJ, **529**, 946–951. (in p. 3)

Predehl, P., Andritschke, R., Böhringer, H., Bornemann, W., Bräuninger, H., Brunner, H., Brusa, M., Burkert, W., Burwitz, V., Cappelluti, N., Churazov, E., Dennerl, K., Eder, J., Elbs, J., Freyberg, M., Friedrich, P., Fürmetz, M., Gaida, R., Hälker, O., Hartner, G., Hasinger, G., Hermann, S., Huber, H., Kendziorra, E., von Kienlin, A., Kink, W., Kreykenbohm, I., Lamer, G., Lapchov, I., Lehmann, K., Meidinger, N., Mican, B., Mohr, J., Mühlegger, M., Müller, S., Nandra, K., Pavlinsky, M., Pfeffermann, E., Reiprich, T., Robrade, J., Rohé, C., Santangelo, A., Schächner, G., Schanz, T., Schmid, C., Schmitt, J., Schreib, R., Schrey, F., Schwope, A., Steinmetz, M., Strüder, L., Sunyaev, R., Tenzer, C., Tiedemann, L., Vongehr, M. & Wilms, J. (2010). eROSITA on SRG. In *Space Telescopes and Instrumentation 2010: Ultraviolet to Gamma Ray*, vol. 7732 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 77320U. (in p. 104)

Prestwich, A.H., Tsantaki, M., Zezas, A., Jackson, F., Roberts, T.P., Foltz, R., Linden, T. & Kalogera, V. (2013). Ultra-luminous X-Ray Sources in the Most Metal Poor Galaxies. ApJ, **769**, 92. (in pp. 11, 65, and 95)

Ptak, A. & Griffiths, R. (1999). Hard X-Ray Variability in M82: Evidence for a Nascent Active Galactic Nucleus? ApJ, **517**, L85–L89. (in p. 63)

Rappaport, S.A., Podsiadlowski, P. & Pfahl, E. (2005). Stellar-mass black hole binaries as ultraluminous X-ray sources. MNRAS, **356**, 401–414. (in pp. 4, 9, 10, and 65)

Raychaudhury, S., Vargas, I.M., Jones, C., Forman, W.R., Vikhlinin, A. & Murray, S.S. (2008). The nature of Ultra-luminous X-ray sources in early-type galaxies. In S.K. Chakrabarti & A.S. Majumdar, eds., *American Institute of Physics Conference Series*, vol. 1053 of *American Institute of Physics Conference Series*, 111–115. (in p. 98)

Reig, P. (2011). Be/X-ray binaries. Ap&SS, **332**, 1–29. (in p. 2)

Remillard, R.A. & McClintock, J.E. (2006). X-Ray Properties of Black-Hole Binaries. ARA&A, **44**, 49–92. (in pp. 9 and 64)

Risaliti, G. & Elvis, M. (2004). *A Panchromatic View of AGN*, vol. 308 of *Astrophysics and Space Science Library*, 187. Barger, A. J. (in p. 110)

Roberts, T.P. & Warwick, R.S. (2000). A ROSAT High Resolution Imager survey of bright nearby galaxies. MNRAS, **315**, 98–114. (in p. 65)

Roberts, T.P., Warwick, R.S., Ward, M.J. & Murray, S.S. (2002). A Chandra observation of the interacting pair of galaxies NGC 4485/4490. MNRAS, **337**, 677–692. (in pp. 11 and 65)

Rodighiero, G., Daddi, E., Baronchelli, I., Cimatti, A., Renzini, A., Aussel, H., Popesso, P., Lutz, D., Andreani, P., Berta, S., Cava, A., Elbaz, D., Feltre, A., Fontana, A., Förster Schreiber, N.M., Franceschini, A., Genzel, R., Grazian, A., Gruppioni, C., Ilbert, O., Le Floch, E., Magdis, G., Magliocchetti, M., Magnelli, B., Maiolino, R., McCracken, H., Nordon, R., Poglitsch, A., Santini, P., Pozzi, F., Riguccini, L., Tacconi, L.J., Wuyts, S. & Zamorani, G. (2011). The Lesser Role of Starbursts in Star Formation at z = 2. ApJ, **739**, L40. (in p. 90)

Rowan-Robinson, M. (1999). The star-formation history of the universe. Ap&SS, **266**, 291–300. (in pp. 35 and 69)

Salim, S., Lee, J.C., Janowiecki, S., da Cunha, E., Dickinson, M., Boquien, M., Burgarella, D., Salzer, J.J. & Charlot, S. (2016). GALEX-SDSS-WISE Legacy Catalog (GSWLC): Star Formation Rates, Stellar Masses, and Dust Attenuations of 700,000 Low-redshift Galaxies. ApJS, **227**, 2. (in pp. 30 and 31)

Salmon, L., Hanlon, L., Jeffrey, R.M. & Martin-Carrillo, A. (2020). Web application for galaxy-targeted follow-up of electromagnetic counterparts to gravitational wave sources. A&A, **634**, A32. (in pp. 43 and 44)

Sanders, D.B., Mazzarella, J.M., Kim, D.C., Surace, J.A. & Soifer, B.T. (2003). The IRAS Revised Bright Galaxy Sample. AJ, **126**, 1607–1664. (in p. 28)

Saxton, R.D., Read, A.M., Esquej, P., Freyberg, M.J., Altieri, B. & Bermejo, D. (2008). The first XMM-Newton slew survey catalogue: XMMSL1. A&A, **480**, 611–622. (in p. 16)

Sazonov, S. & Khabibullin, I. (2018). Impact of ultraluminous X-ray sources on photoabsorption in the first galaxies. MNRAS, **476**, 2530–2541. (in p. 10)

Schutz, B.F. (1986). Determining the Hubble constant from gravitational wave observations. Nature, **323**, 310–311. (in p. 43)

Selsing, J., Krühler, T., Malesani, D., D'Avanzo, P., Schulze, S., Vergani, S.D., Palmerio, J., Japelj, J., Milvang-Jensen, B., Watson, D., Jakobsson, P., Bolmer, J., Cano, Z., Covino, S., D'Elia, V., de Ugarte Postigo, A., Fynbo, J.P.U., Gomboc, A., Heintz, K.E., Kaper, L., Levan, A.J., Piranomonte, S., Pugliese, G., Sánchez-Ramírez, R., Sparre, M., Tanvir, N.R., Thöne, C.C. & Wiersema, K. (2018). The host galaxy of the short GRB 111117A at z = 2.211. Impact on the short GRB redshift distribution and progenitor channels. A&A, **616**, A48. (in p. 48)

Seward, F.D. & Charles, P.A. (2010). *Exploring the X-ray Universe*. Cambridge Univeristy Press, New York. (in p. 2)

Shakura, N.I. & Sunyaev, R.A. (1973). Reprint of 1973A&A....24..337S. Black holes in binary systems. Observational appearance. A&A, **500**, 33–51. (in pp. 9 and 10)

She, R., Ho, L.C. & Feng, H. (2017). Chandra Survey of Nearby Galaxies: The Catalog. ApJ, **835**, 223. (in pp. 36, 62, and 80)

Shklovsky, I.S. (1967). On the Nature of the Source of X-Ray Emission of Sco XR-1. ApJ, **148**, L1. (in p. 1)

Singer, L.P., Chen, H.Y., Holz, D.E., Farr, W.M., Price, L.R., Raymond, V., Cenko, S.B., Gehrels, N., Cannizzo, J., Kasliwal, M.M., Nissanke, S., Coughlin, M., Farr, B., Urban, A.L., Vitale, S., Veitch, J., Graff, P., Berry, C.P.L., Mohapatra, S. & Mandel, I. (2016). Going the Distance: Mapping Host Galaxies of LIGO and Virgo Sources in Three Dimensions Using Local Cosmography and Targeted Follow-up. ApJ, **829**, L15. (in p. 44)

Sądowski, A., Narayan, R., McKinney, J.C. & Tchekhovskoy, A. (2014a). Numerical simulations of super-critical black hole accretion flows in general relativity. MNRAS, **439**, 503–520. (in p. 10)

Sądowski, A., Narayan, R., McKinney, J.C. & Tchekhovskoy, A. (2014b). Numerical simulations of super-critical black hole accretion flows in general relativity. MNRAS, **439**, 503–520. (in p. 64)

Skrutskie, M.F., Cutri, R.M., Stiening, R., Weinberg, M.D., Schneider, S., Carpenter, J.M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D.G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J.D., Gizis, J.E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E.L., Marsh, K.A., McCallon, H.L., Tam, R., Van Dyk, S. & Wheelock, S. (2006). The Two Micron All Sky Survey (2MASS). AJ, **131**, 1163–1183. (in p. 29)

Soria, R., Cropper, M., Pakull, M., Mushotzky, R. & Wu, K. (2005). The star-forming environment of an ultraluminous X-ray source in NGC4559: an optical study. MNRAS, **356**, 12–28. (in pp. 11, 65, and 95)

Soria, R., Kuntz, K.D., Winkler, P.F., Blair, W.P., Long, K.S., Plucinsky, P.P. & Whitmore, B.C. (2012). The Birth of an Ultraluminous X-Ray Source in M83. ApJ, **750**, 152. (in p. 126)

Speagle, J.S., Steinhardt, C.L., Capak, P.L. & Silverman, J.D. (2014). A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6. ApJS, **214**, 15. (in p. 90)

Springob, C.M., Haynes, M.P., Giovanelli, R. & Kent, B.R. (2005). A Digital Archive of H I 21 Centimeter Line Spectra of Optically Targeted Galaxies. ApJS, **160**, 149–162. (in p. 56)

Stampoulis, V., van Dyk, D.A., Kashyap, V.L. & Zezas, A. (2019). Multidimensional data-driven classification of emission-line galaxies. MNRAS, **485**, 1085–1102. (in pp. 36, 70, and 80)

Steer, I., Madore, B.F., Mazzarella, J.M., Schmitz, M., Corwin, H.G., Chan, B.H.P., Ebert, R., Helou, G., Baker, K., Chen, X., Frayer, C., Jacobson, J., Lo, T., Ogle, P., Pevunova, O. & Terek, S. (2017). Redshift-independent Distances in the NASA/IPAC Extragalactic Database: Methodology, Content, and Use of NED-D. AJ, **153**, 37. (in pp. 19, 20, and 69)

Sun, N.C., de Grijs, R., Cioni, M.R.L., Rubele, S., Subramanian, S., van Loon, J.T., Bekki, K., Bell, C.P.M., Ivanov, V.D., Marconi, M., Muraveva, T., Oliveira, J.M. & Ripepi, V. (2018). The VMC Survey. XXIX. Turbulence-controlled Hierarchical Star Formation in the Small Magellanic Cloud. ApJ, **858**, 31. (in p. 73)

Svoboda, J., Douna, V., Orlitová, I. & Ehle, M. (2019). Green Peas in X-Rays. ApJ, **880**, 144. (in p. 97)

Swartz, D.A., Ghosh, K.K., Tennant, A.F. & Wu, K. (2004). The Ultraluminous X-Ray Source Population from the Chandra Archive of Galaxies. ApJS, **154**, 519–539. (in pp. 11, 64, 65, 83, 85, and 98)

Swartz, D.A., Soria, R. & Tennant, A.F. (2008). Do Ultraluminous X-Ray Sources Exist in Dwarf Galaxies? ApJ, **684**, 282–286. (in pp. 11, 65, and 95)

Swartz, D.A., Soria, R., Tennant, A.F. & Yukita, M. (2011). A Complete Sample of Ultraluminous X-ray Source Host Galaxies. ApJ, **741**, 49. (in pp. 11, 66, 70, 77, 83, 84, 92, 110, 111, 119, and 137)

Tanaka, Y. (1989). Black holes in X ray binaries: X ray properties of the galactic black hole candidates. In J. Hunt & B. Battrick, eds., *Two Topics in X-Ray Astronomy, Volume 1: X Ray Binaries. Volume 2: AGN and the X Ray Background*, vol. 296 of *ESA Special Publication*. (in p. 9)

Tanvir, N.R., Levan, A.J., Fruchter, A.S., Hjorth, J., Hounsell, R.A., Wiersema, K. & Tunnicliffe, R.L. (2013). A 'kilonova' associated with the short-duration γ-ray burst GRB 130603B. Nature, **500**, 547–549. (in p. 48)

Tauris, T.M. & van den Heuvel, E.P.J. (2006). Formation and evolution of compact stellar X-ray sources. In W.H.G. Lewin & M. van der Klis, eds., *Compact stellar X-ray sources*, 623–665, Cambridge University Press. (in p. 5)

Terry, J.N., Paturel, G. & Ekholm, T. (2002). Local velocity field from sosie galaxies. I. The Peebles' model. A&A, **393**, 57–68. (in p. 53)

Toffano, M., Mapelli, M., Giacobbo, N., Artale, M.C. & Ghirlanda, G. (2019). The host galaxies of double compact objects across cosmic time. MNRAS, **489**, 4622–4631. (in p. 43)

Tremonti, C.A., Heckman, T.M., Kauffmann, G., Brinchmann, J., Charlot, S., White, S.D.M., Seibert, M., Peng, E.W., Schlegel, D.J., Uomoto, A., Fukugita, M. & Brinkmann, J. (2004). The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey. ApJ, **613**, 898–913. (in pp. 15, 30, and 70)

Tully, R.B., Courtois, H.M. & Sorce, J.G. (2016). Cosmicflows-3. AJ, **152**, 50. (in pp. 19 and 22)

Tzanavaris, P., Hornschemeier, A.E., Gallagher, S.C., Lenkić, L., Desjardins, T.D., Walker, L.M., Johnson, K.E. & Mulchaey, J.S. (2016). Exploring X-Ray Binary Populations in Compact Group Galaxies with Chandra. ApJ, **817**, 95. (in pp. 11, 65, and 95)

van der Marel, R.P. (2004). Intermediate-mass Black Holes in the Universe: A Review of Formation Theories and Observational Constraints. *Coevolution of Black Holes and Galaxies*, 37. (in pp. 9 and 63)

Venkatesan, A., Giroux, M.L. & Shull, J.M. (2001). Heating and Ionization of the Intergalactic Medium by an Early X-Ray Background. ApJ, **563**, 1–8. (in p. 64)

Verdes-Montenegro, L., Sulentic, J., Lisenfeld, U., Leon, S., Espada, D., Garcia, E., Sabater, J. & Verley, S. (2005). The AMIGA project. I. Optical characterization of the CIG catalog. A&A, **436**, 443–455. (in p. 56)

Walton, D.J., Roberts, T.P., Mateos, S. & Heard, V. (2011). 2XMM ultraluminous X-ray source candidates in nearby galaxies. MNRAS, **416**, 1844–1861. (in pp. 11, 65, 98, and 99)

Walton, D.J., Nardini, E., Gallo, L.C., Reynolds, M.T., Ricci, C., Dauser, T., Fabian, A.C., García, J.A., Harrison, F.A., Risaliti, G. & Stern, D. (2019). A low-flux state in IRAS 00521-7054 seen with NuSTAR and XMM-Newton: relativistic reflection and an ultrafast outflow. MNRAS, **484**, 2544–2555. (in p. 64)

Wang, L., Rowan-Robinson, M., Norberg, P., Heinis, S. & Han, J. (2014). The Revised IRAS-FSC Redshift Catalogue (RIFSCz). MNRAS, **442**, 2739–2750. (in p. 28)

Wang, S., Qiu, Y., Liu, J. & Bregman, J.N. (2016). Chandra ACIS Survey of X-Ray Point Sources in Nearby Galaxies. II. X-Ray Luminosity Functions and Ultraluminous X-Ray Sources. ApJ, **829**, 20. (in pp. 11, 66, 70, 76, 78, 79, 80, 83, 84, 85, 86, 92, 99, 110, 111, 112, 113, and 119)

Webb, N.A., Coriat, M., Traulsen, I., Ballet, J., Motch, C., Carrera, F.J., Koliopanos, F., Authier, J., de la Calle, I., Ceballos, M.T., Colomo, E., Chuard, D., Freyberg, M., Garcia, T., Kolehmainen, M., Lamer, G., Lin, D., Maggi, P., Michel, L., Page, C.G., Page, M.J., Perea-Calderon, J.V., Pineau, F.X., Rodriguez, P., Rosen, S.R., Santos Lleo, M., Saxton, R.D., Schwope, A., Tomás, L., Watson, M.G. & Zakardjian, A. (2020). The XMM-Newton serendipitous survey. IX. The fourth XMM-Newton serendipitous source catalogue. A&A, **641**, A136. (in p. 37)

Wenger, M., Ochsenbein, F., Egret, D., Dubois, P., Bonnarel, F., Borde, S., Genova, F., Jasniewicz, G., Laloë, S., Lesteven, S. & Monier, R. (2000). The SIMBAD astronomical database. The CDS reference database for astronomical objects. A&AS, **143**, 9–22. (in p. 15)

White, D.J., Daw, E.J. & Dhillon, V.S. (2011). A list of galaxies for gravitational wave searches. *Classical and Quantum Gravity*, **28**, 085016. (in pp. 16, 38, and 43)

Wiktorowicz, G., Sobolewska, M., Lasota, J.P. & Belczynski, K. (2017). The Origin of the Ultraluminous X-Ray Sources. ApJ, **846**, 17. (in pp. 9, 11, 65, 93, 99, 100, 101, 102, and 137)

Wolf, C., Onken, C.A., Luvaul, L.C., Schmidt, B.P., Bessell, M.S., Chang, S.W., Da Costa, G.S., Mackey, D., Martin-Jones, T., Murphy, S.J., Preston, T., Scalzo, R.A., Shao, L., Smillie, J., Tisserand, P., White, M.C. & Yuan, F. (2018). SkyMapper Southern Survey: First Data Release (DR1). PASA, **35**, e010. (in p. 56)

Wolter, A. & Trinchieri, G. (2004). A thorough study of the intriguing X-ray emission from the Cartwheel ring. A&A, **426**, 787–796. (in pp. 11 and 65)

Wolter, A., Fruscione, A. & Mapelli, M. (2018). The X-Ray Luminosity Function of Ultraluminous X-Ray Sources in Collisional Ring Galaxies. ApJ, **863**, 43. (in p. 65)

Wyatt, S.D., Tohuvavohu, A., Arcavi, I., Lundquist, M.J., Howell, D.A. & Sand, D.J. (2020). The Gravitational Wave Treasure Map: A Tool to Coordinate, Visualize, and Assess the Electromagnetic Follow-up of Gravitational-wave Events. ApJ, **894**, 127. (in p. 43)

Yang, S., Sand, D.J., Valenti, S., Cappellaro, E., Tartaglia, L., Wyatt, S., Corsi, A., Reichart, D.E., Haislip, J., Kouprianov, V. & (DLT40 Collaboration (2019). Optical Follow-up of Gravitational-wave Events during the Second Advanced LIGO/VIRGO Observing Run with the DLT40 Survey. ApJ, **875**, 59. (in pp. 43 and 45)

Zevin, M., Kelley, L.Z., Nugent, A., Fong, W.f., Berry, C.P.L. & Kalogera, V. (2019). Forward Modeling of Double Neutron Stars: Insights from Highly-Offset Short Gamma-Ray Bursts. *arXiv e-prints*, arXiv:1910.03598. (in p. 48)

Zezas, A. & Fabbiano, G. (2002). Chandra Observations of "The Antennae" Galaxies (NGC 4038/4039). IV. The X-Ray Source Luminosity Function and the Nature of Ultraluminous X-Ray Sources. ApJ, **577**, 726–737. (in p. 9)

Zezas, A., Ward, M.J. & Murray, S.S. (2003). Chandra Observations of the X-Ray-luminous Star-forming Galaxy Merger Arp 299. ApJ, **594**, L31–L34. (in pp. 11 and 98)

Zezas, A., Fabbiano, G., Baldi, A., Schweizer, F., King, A.R., Rots, A.H. & Ponman, T.J. (2007). Chandra Monitoring Observations of the Antennae Galaxies. II. X-Ray Luminosity Functions. ApJ, **661**, 135–148. (in pp. 11, 65, and 76)

Zhang, Z., Gilfanov, M. & Bogdán, Á. (2012). Dependence of the low-mass X-ray binary population on stellar age. A&A, **546**, A36. (in pp. 5, 6, 7, 8, 11, 98, 99, 100, 116, and 137)