The stellar content of soft X-ray surveys

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Sebastian Freund

aus Uelzen

Hamburg

2022

Gutachter der Dissertation: Prof. Dr. Jürgen H. M. M. Schmitt
Prof. Dr. Manuel Güdel

Zusammensetzung der Prüfungkommission: Prof. Dr. Jochen Liske
Prof. Dr. Jürgen H. M. M. Schmitt
Prof. Dr. Manuel Güdel
Prof. Dr. Marcus Brüggen
Prof. Dr. Peter H. Hauschildt

Vorsitzender der Prüfungskomission: Prof. Dr. Jochen Liske

Datum der Disputation: 07.07.2022

Vorsitzender Fach-Promotionsausschusses PHYSIK: Prof. Dr. Wolfgang J. Parak

Leiter des Fachbereichs PHYSIK: Prof. Dr. Günter H. W. Sigl

Dekan der Fakultät MIN: Prof. Dr. Heinrich Graener

Zusammenfassung

In der Chromosphäre kühler Sterne beginnt die Temperatur mit zunehmender Höhe zu steigen und in der Korona werden Werte von mehr als 10^6 K erreicht. Daher emittiert die Korona den größten Teil ihrer Leuchtkraft als weiche Röntgenstrahlung. Ihr Beitrag zur Gesamtenergieabgabe des Sterns ist aufgrund der geringen Dichte der Korona jedoch gering. Der Mechanismus, der zur Aufheizung der Korona führt, ist noch nicht vollständig geklärt, es wird jedoch angenommen, dass Magnetfelder eine wesentliche Rolle spielen. Die Untersuchung großer Samples stellarer Röntgenquellen ermöglicht es, die Beziehung zwischen der Röntgenaktivität und der Masse, dem Alter und der Sternrotation zu bestimmen.

Counterparts im optischen oder nahen Infrarot (NIR) zu den Röntgenquellen liefern zusätzliche Informationen über die Quelle, z.B. optische Helligkeit, Farbe und Parallaxe, und helfen bei der Identifikation von stellaren Quellen. Die Wahrscheinlichkeit, Counterparts zu erhalten, die sich zufällig in der gleichen Richtung befinden, aber nicht mit der Röntgenquelle assoziiert sind, ist jedoch oft hoch, da die Positionsunsicherheit von Röntgendetektionen im Vergleich zu anderen Wellenlängenbereichen meistens groß ist.

Diese Arbeit befasst sich mit Methoden zur Identifizierung des stellaren Anteils von großen Röntgendurchmusterungen. Aufgrund der großen Anzahl von Röntgenquellen werden automatische Algorithmen vorgestellt, die typischerweise geometrische und weitere Eigenschaften der Counterparts berücksichtigen. Die daraus resultierenden flusslimitierten Samples von stellaren Röntgenquellen werden anschließend analysiert.

Nach einer Einführung in den Hintergrund dieser Arbeit stelle ich den stellaren Anteil des *XMM-Newton* Slew Surveys (XMMSL2) vor, der ausschließlich aus den geometrischen Abständen zwischen den XMMSL2-Quellen und Counterparts aus verschiedenen optischen und NIR-Katalogen bestimmt wird. Das Ergebnis besteht aus 5 920 glaubwürdigen stellare Röntgenquellen und es wird eine Vollständigkeit und Zuverlässigkeit von mehr als 96 % erwartet.

Für die erste Identifizierung des gesamten stellaren Anteils des *ROSAT* All-Sky Surveys (RASS) werden Bayes-Maps verwendet, um weitere Eigenschaften der Counterpart zusätzlich zur geometrischen Übereinstimmung zu berücksichtigen. Es wird gezeigt, dass die Ergebnisse nur geringfügig von den Details der Schätzung der Bayes-Map abhängen. Mit diesem Verfahren werden mehr als 28 000 RASS-Quellen als stellar identifiziert, was die bisher größte Auswahl stellarer Röntgenquellen darstellt. Es handelt sich um ein flusslimitiertes Sample, dass alle Typen von Sternen enthält, die zuvor separat untersucht wurden. Anhand der für jede Quelle angegebenen Wahrscheinlichkeiten wird die Vollständigkeit und Zuverlässigkeit auf etwa 93 % geschätzt.

Ein ähnliches Identifikationsverfahren wird auch auf den *eROSITA* Final Equatorial-Depth Survey (eFEDS) angewandt und die Ergebnisse werden mit der Identifizierung durch eine maschinelle Lernmethode und einen anderen Bayes'schen Ansatz verglichen, der nicht auf Sterne spezialisiert ist. Von den eFEDS-Detektionen werden 2 060 Quellen als Sterne identifiziert mit einer Vollständigkeit und Zuverlässigkeit von etwa 90 %. Die Identifikationen werden durch die anderen Verfahren ebenfalls zu etwa 90 % bestätigt.

Die stellaren Identifikationen von RASS und gezielten Beobachtungen mit *ROSAT*, *Chandra* und *XMM-Newton* werden zur Analyse der Röntgeneigenschaften der Hyaden verwendet. Für 281 von 1066 Hyaden-Mitgliedern werden Röntgendetektionen gefunden. Die Röntgenaktivität nimmt für Quellen von späteren Spektraltypen zu und erreicht die Sättigungsgrenze etwa bei M3-Zwergen. Die Rotations-Aktivitäts-Relation der Hyaden-Mitglieder bestätigt frühere Ergebnisse.

Mit dem in dieser Arbeit vorgestellten Identifikationsverfahren und einigen Modifikationen wird es möglich sein, den stellaren Inhalt des derzeit laufenden *eROSITA* All-Sky Surveys (eRASS) zu identifizieren, was unser Verständnis der stellaren Aktivität revolutionieren wird.

Abstract

In the chromosphere of cool stars, the temperature starts rising with increasing height and values of more than 10^6 K are reached in the corona. Hence, the corona emits most of its luminosity in soft X-rays. Its contribution to the total energy output of the star is small due to the low density of the corona. The mechanism of the coronal heating is still not fully understood but magnetic fields are thought to play a fundamental role. The study of large samples of stellar X-ray sources allows to determine the relation of the X-ray activity on mass, age, and stellar rotation.

Counterparts in the optical or near infrared (NIR) to the X-ray sources provide additional information about the source, e.g., optical magnitude, color, and parallax, and help to identify stellar sources. However, the probability of obtaining counterparts that are by chance located in the same direction but are not associated to the X-ray source is often high due to large positional uncertainties of X-ray detections compared to other wavelength regimes.

This thesis deals with methods to identify the stellar content of large X-ray surveys. Due to the high number of X-ray sources, automatic algorithms are presented that typically take into account geometric and further properties of the counterparts. The resulting flux-limited samples of stellar X-ray sources are subsequently analyzed.

After an introduction to the background of this thesis, I present the stellar content of the *XMM-Newton* slew survey (XMMSL2) that is solely obtained by the geometric separations between the XMMSL2 sources and counterpart from various optical and NIR catalogs. The obtained sample contains 5 920 bona fide stellar X-ray emitters and is expected to be complete and reliable to more than 96 %.

For the first identification of the whole stellar content of the *ROSAT* all-sky survey (RASS), Bayes maps are applied to consider further properties of the counterparts in addition to the geometric match. It is shown that the results only weakly depend on the details of the estimation of the Bayes map. With this procedure, more than 28 000 RASS sources are identified as stellar, which is the largest sample of stellar X-ray sources presented so far. it is a flux-limited sample containing all types of stars that were previously investigated separately. From matching probabilities provided for every source, the completeness and reliability is estimated to be about 93 %.

A similar identification procedure is also applied to the *eROSITA* Final Equatorial-Depth Survey (eFEDS) and the results are compared to the identification of a machine learning method and another Bayesian approach that is not specialized to stars. Of the eFEDS detections, 2 060 sources are identified as stars with a completeness and reliability of about 90 %. The identifications are confirmed by the other procedures also to about 90 %.

The stellar identifications from the RASS and pointing observations with *ROSAT*, *Chandra*, and *XMM-Newton* are applied to analyze the X-ray properties of the Hyades cluster. X-ray detections for 281 of 1066 Hyades members are found. The X-ray activities increase for sources of later spectral types reaching the saturation limit for about M3 dwarfs. The rotation-activity relation of the Hyades members confirms previous results.

With the identification procedure presented in this thesis and some modifications, it will be possible to identify the stellar content of the currently ongoing *eROSITA* all-sky survey (eRASS) and to revolutionize our understanding of stellar activity.

Contents

Chapter 1

Introduction

Only due to the coincidence that, in our time, the Moon apparently covers an area of nearly the same size in the sky as the Sun, humans in all times could admire the solar corona in scattered visible light during solar eclipses (see Fig. 1.1). Outside the few minutes of a solar eclipse, when the Moon blocks the light of the photosphere, the corona is greatly outshone at optical wavelengths because of its much lower density. About 80 years ago, Grotrian (1939) and Edlén (1943) explained optical coronal lines by the emission of highly ionized atoms whose existence requires very high temperatures. This hot plasma emits most of its radiation in the X-ray regime, however, X-ray radiation is absorbed by Earth's atmosphere and cannot be detected from ground. Therefore, coronal X-ray emission could not be observed until rockets equipped with Geiger counters were launched (Burnight, 1949). Later, coronal emission from stars was detected during rocket flights with more sensitive X-ray telescopes (Catura et al., 1975). These first stellar X-ray sources exhibited much stronger X-ray emission compared to their optical flux than the Sun.

Satellites dedicated to X-ray astronomy revealed that coronal X-ray emission can be found for nearly all stars with an outer convection zone. The X-ray activity is mainly related to stellar rotation, and hence, to age (Skumanich, 1972). For very young and fast rotating stars, the X-ray flux saturates as a function of the bolometric flux. Magnetic fields seem to play a fundamental role in the heating of the corona (Pevtsov et al., 2003), but the details are still not fully understood. Although stars emit only a small fraction of their luminosity in X-rays, the high energy radiation and the stellar magnetic fields have a large impact on the evolution of accretion of protostars, planet formation, and the chemical development of planetary atmospheres.

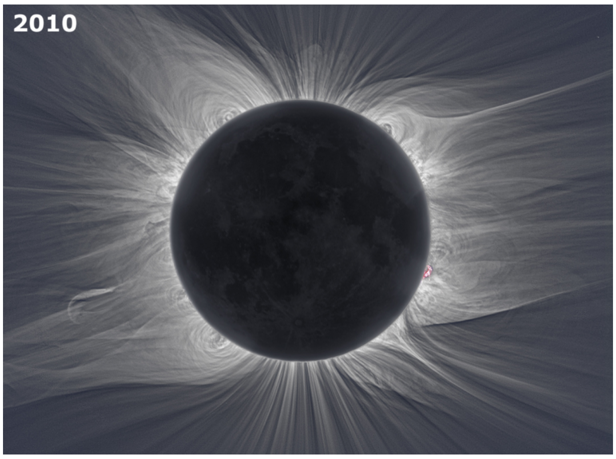


Figure 1.1: White light image of the 2010 solar eclipse. Adopted from Fig. 4 in Habbal et al. (2021).

Before the properties of the X-ray sources can be investigated, the source type has to be classified. Since the optical flux of stellar sources is much higher than the X-ray flux, optical counterparts to the coronal X-ray sources are generally detectable and provide additional information, e.g., parallax, spectral type, and luminosity class. However, the identification of the correct counterpart is often challenging because the positional uncertainties of X-ray sources are generally much larger than for optical sources. Therefore, often multiple potential counterparts are located in the error circle. For the early X-ray missions, spectra of all counterparts in the searching area were taken to find reasonable identifications (Gioia et al., 1990; Stocke et al., 1991), but this became unfeasible as the size of the X-ray catalogs grew. Then, X-ray counterparts to selected objects of interest were searched for, based on the positions of the sources. However, the obtained samples are not complete neither according to their optical nor X-ray flux. Since the identifications of the various source types differ by their completeness and reliability, the sample sizes and the X-ray properties of the stellar source classes are not comparable. Therefore, flux-limited samples of stellar X-ray sources are necessary. The assessment of the full, flux-limited stellar content of the X-ray surveys requires the consideration of additional source properties and a homogeneous optical catalog of sufficient depth.

Such a catalog is provided by the *Gaia* mission (Gaia Collaboration et al., 2016b) that was launched in December 2013. During its nominal five year mission lifetime, all-sky measurements of highly accurate positions, parallaxes, and proper motions for more than 1 billion stars are obtained. Furthermore, photometric measurements in three different bands, radial velocities, object classifications, and stellar parameters obtained from spectro-photometric and spectral measurements are provided. Due to the long mission lifetime, intermediate data releases are regularly published that are constructed from an increasing amount of input data, improved calibration and data reduction. They provide more parameters with increasing accuracies.

On July 13th, 2019 a new X-ray telescope, the extended ROentgen Survey with an Imaging Telescope Array (eROSITA; Predehl et al., 2021) was launched on board of the German-Russian Spectrum-Roentgen-Gamma (SRG) mission. The main objective of the *eROSITA* mission is to perform an all-sky survey of unprecedented depth in the soft X-ray regime and the first imaging all-sky survey above 2.4 keV. During this survey, about 0.8 million stellar sources are expected to be detected (Merloni et al., 2012). A method to identify the stellar content of this survey is essential to harvest its full scientific potential.

This thesis deals with identification procedures of stellar X-ray sources, its results, and applications. In Chapter 2 I outline the background of this work. Specifically, I introduce the X-ray telescopes, especially those that are most import in the context of this thesis, and describe the properties of coronal X-ray sources. I outline the Gaia mission that provide most of the stellar counterparts in my publications. Then, I summarize previous methods to identify stellar X-ray sources and introduce the identification procedure that is based on Bayesian statistics. When the positional accuracy of the X-ray sources is high and the counterpart density is low, the correct associations can be reliably identified by the geometric match. This is demonstrated in Chapter 3 for the *XMM-Newton* slew survey. However, further properties need to be considered when several potential counterpart are located in the matching circle because of a high positional uncertainty or counterpart density. In Chapter 4 Bayes maps are applied to identify the stellar content of the *ROSAT* all-sky survey. Furthermore, the identification procedure is validated by applying it to a test catalog with known source types, and the construction and influence of the Bayes map is discussed. Then, the identification procedure is applied to the *eROSITA* Final Equatorial-Depth Survey and the results are compared to other identification methods in Chapter 5. In Chapter 6 I apply the stellar identifications to update the X-ray view of the Hyades cluster. Finally, I draw my conclusions and provide an outlook to future applications and improvements in Chapter 7.

This thesis is a cumulative work and the presented publications are sorted according to their content and not chronologically. I described my contributions to the publications in the according chapters.

Chapter 2

Stellar X-ray emission and its optical counterparts

2.1 X-ray telescopes

Compared to other wavelength, e.g. optical, infrared, or radio, astronomy in the X-ray regime started relatively late in 1949 (Burnight, 1949). However, its sensitivity increased in 36 years by a factor of 100 million, which is a similar improvement as the *Hubble Space Telescope* compared to the telescope of Galileo Galilei[1]. In this chapter, I briefly outline the timeline of X-ray telescopes dedicated for observations outside of the solar system in Sect. 2.1.1. In Sects. 2.1.2, 2.1.3, and 2.1.4, I describe the *ROSAT*, *XMM-Newton*, and *eROSITA* telescopes and missions, which are the most important X-ray telescopes to this work.

2.1.1 A brief timeline

After the first X-ray observation of the Sun (Burnight, 1949), further rocket based observations were performed and the low-mass X-ray binary Sco X-1 was detected in 1962 as the first X-ray source outside the solar system (Giacconi et al., 1962). However, in X-ray binaries, the X-ray emission is produced by matter from a star falling on a compact object and not by the stellar corona as for the Sun (see Sect. 2.2.1). In December 1970, the first satellite dedicated to X-ray astronomy, *Uhuru*, was launched into Earth orbit (Giacconi et al., 1971). It performed the first all-sky survey and detected 339 X-ray sources but still no coronal stellar emission had been observed.

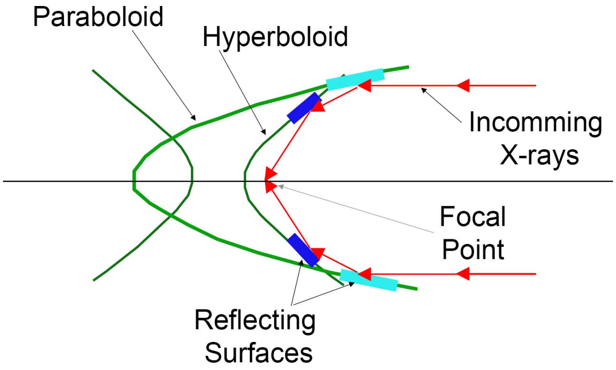


Figure 2.1: Wolter type I grazing incidence telescope. Credit: NASA's image the Universe (`https://imagine.gsfc.nasa.gov/observatories/technology/xray_telescopes2.html`).

These first rockets carried Geiger counters, however, to achieve sensitivities that are necessary to observe stellar X-ray emission, the X-rays need to be focused. Conventional telescope designs used for optical or infrared light do not work for X-rays because they are not reflected but instead transmitted or absorbed when striking the mirror surface nearly perpendicular. Thus, Wolter (1952) proposed a design with grazing incidence mirrors originally for an X-ray microscope that was later adopted for an X-ray telescope. In the so-called Wolter type I telescopes, the X-rays are reflected two times on a parabolic and hyperbolic shaped metal surface as shown in Fig. 2.1. The effective area can be further increased by folding several mirror shells within each other. Applying a similar telescope,

---

[1] `https://chandra.harvard.edu/xray_astro/history.html`

Catura et al. (1975) identified the RS CVn system Capella as the first stellar X-ray source.

A very important mission for X-ray astronomy was the *Einstein* Observatory (Giacconi et al., 1979) launched in 1978. As the first fully imaging X-ray satellite, *Einstein* was equipped with a Wolter type I grazing incidence telescope and different instruments mounted on a carousel that could be rotated into the focal plane. The main instruments were the Imaging Proportional Counter (IPC) covering the entire field of view of 75 arcmin with moderate spectral and spatial resolution of about 1 arcmin and the High Resolution Imaging camera (HRI) providing a higher angular resolution of ~ 2 arcsec but covering only the central 25 arcmin of the focal plane. During its nearly 2.5 years of operation, the *Einstein* Observatory performed more than 5000 pointed observations that lifted X-ray astronomy into the mainstream of astronomical research. *Einstein* showed that X-ray emission is found from all stars across the Hertzsprung-Russell diagram (Vaiana et al., 1981).

In June 1990 the *ROSAT* X-ray observatory (Trümper, 1984) was launched and performed the first all-sky survey with an imaging X-ray telescope and more than 9000 pointed observations during 8.5 years of successful operation (see Sect. 2.1.2).

Currently, there are two major X-ray missions, both launched in 1999. The X-ray Multi-Mirror Mission *XMM-Newton* (Jansen et al., 2001) is operated by the European Space Agency (ESA) and focus on a high effective area. It is equipped with three X-ray telescopes that observe the same field of view and consist of 58 nested Wolter type I mirrors (see Sect. 2.1.3). The *Chandra* X-ray Observatory (Weisskopf et al., 2002) by the National Aeronautics and Space Administration (NASA) has a high spatial resolution of 0.5 arcsec. *Chandra* has four nested mirrors and contains as focal plane instruments a High Resolution Camera (HRC) for high resolution imaging and fast timing measurements and an Advanced CCD Imaging Spectrometer (ACIS) that does simultaneous imaging and spectroscopy. High resolution spectra are obtained by the Low and High Energy Transmission Grating (LETG and HETG). Both, *XMM-Newton* and *Chandra*, are dedicated to perform pointed observations of selected targets and the usually do not perform surveys. However, when *XMM-Newton* slews from one pointed observation to the next, its cameras are not switched off but are still collecting data. These data are the basis of the *XMM-Newton* slew survey catalog (XMMSL; Saxton et al., 2008).

In 2019 the extended ROentgen Survey with an Imaging Telescope Array (eROSITA; Predehl et al., 2021) was launched on-board the Spectrum-Roentgen-Gamma (SRG) mission and started a four year all-sky survey (see Sect. 2.1.4).

2.1.2 ROSAT

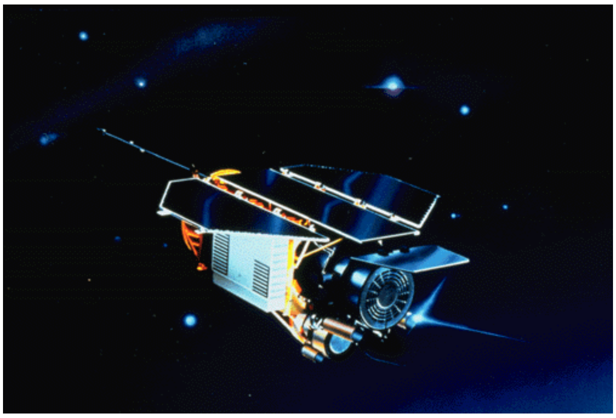


Figure 2.2: X-ray observatory *ROSAT*. Adopted from `https://www2011.mpe.mpg.de/xray/wave/rosat/`

The following description is based on Briel et al. (1996) and further information about the *ROSAT* satellite can be found therein.

The X-ray observatory *ROSAT* was launched in a circular orbit with an altitude of 580 km on June 1$^{\text{st}}$, 1990. An artist's impression of the *ROSAT* satellite is shown in Fig. 2.2. *ROSAT* was designed and built by Germany with contributions from NASA and British Science and Engineering Research Council (SERC), the prime *ROSAT* X-ray data center was at the Max-Planck-Institute for Extraterrestrial Physics (MPE) in Garching.

The X-ray telescope of *ROSAT* consisted of four nested Wolter type I mirrors constructed of Zerodur, a glass ceramic with an almost negligible thermal coefficient, and coated with a thin layer of gold to enhance the X-ray reflectivity. The telescope had a field of view of 2°. In the focal plane, two redundant position-sensitive proportional counters (PSPC) were mounted on a carousel. The PSPCs were multiwire proportional counters that provided a modest energy resolution and a spatial resolution of about 20 arcsec, they were filled with a gas mixture of argon, xenon, and methane. The gas had to be consistently replaced, and thus, it was exhausted in September 1994 and the PSPCs could no longer be used. The thin plastic entrance window was supported against the gas pressure by X-ray opaque ribs in a "wagon wheel" structure and two wire mesh systems.

As an alternative to the PSPCs, the *ROSAT* High Resolution Imager (HRI) could be used that consisted of two cascaded microchannel plates with a crossed grid position readout system. The HRI had a high positional accuracy of about 5 arcsec but a square field of view of only 38 arcmin. The X-ray telescope and the detectors resulted in a high sensitivity between 0.1 and 2.4 keV. In addition to the X-ray telescope, *ROSAT* was equipped with a Wide-Field Camera (WFC) for observations in the EUV (0.06-0.2 keV) range.

One of the main scientific objectives of the *ROSAT* mission was to perform the first all-sky survey with an imaging X-ray telescope. After a verification phase including a mini-survey in July 1990, the *ROSAT* all-sky survey (RASS) started in August 1990 and lasted to January 1991. For the RASS, the PSPC was used because of its large field of view. During the all-sky survey, *ROSAT* rotated with a period of 96 min leading to a scan of a two-degrees-wide strip along a great circle over the Ecliptic poles. Due to the rotation of the Earth and the satellite around the Sun, the scanning strip shifted by about 1° per day so that the survey of the whole sky was completed within half a year. However, due to periods of high background or poor attitude values, some data could not be used and these parts of the sky needed to be re-observed in February and August 1991 and later in February 1997. Since all scanning strips overlap at the Ecliptic poles, the exposure time is here the highest with about 40 000 s and decreases towards the equator to about 400 s. The source catalog of the RASS was first released by (Voges et al., 1999) and later revised by (Boller et al., 2016) adopting an improved data reduction. The first identification of the full stellar content of the RASS is presented by Freund et al., submitted (see Sect. 4.1). After the all-sky survey, *ROSAT* performed pointed observations until it was switched off in February 1999.

2.1.3 XMM-Newton

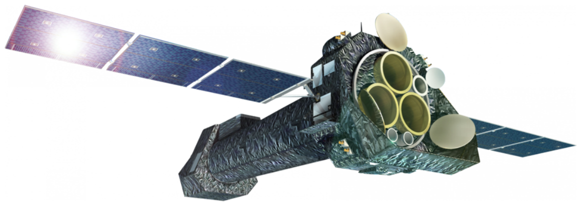


Figure 2.3: X-ray observatory *XMM-Newton*. Adopted from `https://science.nasa.gov/toolkits/spacecraft-icons`

The *XMM-Newton* telescope by ESA was launched in December 1999 and is still observing. In Fig. 2.3 an artist's impression of *XMM-Newton* is shown. The orbit of *XMM-Newton* is highly elliptical with a perigee and apogee of about 7 000 and almost 114 000 km. The resulting orbit period of about 48 hours allows long uninterrupted observations (Jansen et al., 2001). The three telescopes achieve a high collecting area between roughly 150 eV and 12 keV and cover a field of view of 30 arcmin. The on-axis width of the point spread function is between 4.2 and 6.6 arcsec.

In the focal planes of the X-ray telescopes three European Photon Imaging Cameras (EPIC) of two different kinds are mounted. Two of the telescopes are equipped with a Metal Oxide Semi-conductor (MOS) camera (Turner et al., 2001) that consists of seven front-illuminated charge-coupled devices (CCD). The focal point is located on the central CCD and the outer CCDs slightly overlap to reduce the dead regions and they are located 4.5 mm closer to the mirror to better follow the curvature of the focal plane. The full MOS camera contains 600 × 600 pixels and each pixel covers 1.1 arcsec. It is sensitive in the energy range of 0.2 to 10 keV but the quantum efficiency considerably drops below $\sim$ 700 eV. The pn camera (Strüder et al., 2001) that is installed in the focal plane of the third telescope was especially designed for *XMM-Newton* and covers 97 % of the field of view. It is illuminated from the rear side, and for redundancy reasons, the detector is divided in twelve CCDs with 200 × 64 pixels each that are mounted as monolithic fabrication on a single wafer and cover 4.1 arcsec of the field of view. The pn camera achieves high quantum efficiencies over the whole energy range of *XMM-Newton* that only slightly decreases below 600 eV and above 10 keV. Both cameras can be operated in different modes that either allow to cover a large field of view or to improve the time resolution.

Since the EPIC detectors are also sensitive to IR, visible and UV photons, blocking filters of three different thicknesses can be chosen for the observations. The filters reduce the problem of the so called optical loading for optically bright sources but they also reduce the effective area, especially for the detection of soft X-rays. The EPIC cameras provide a moderate spectral resolution of $E/\Delta E \approx 20 - 50$, higher spectral resolutions are obtained by the Reflection Grating Spectrometers (RGS; den Herder et al., 2001). About half of the light of the X-ray telescopes that are equipped with a MOS camera is diverted to the RGS that diffract the X-rays to an array of CCDs. With the RGS, a spectral resolution of 150 to 800 in the energy range of 0.33 to 2.5 keV is obtained.

The main objective of *XMM-Newton* is to perform pointed observations of selected targets. From the datasets of the observations that were publicly available by December 2020, the 4XMM-DR11 catalog is constructed (Traulsen et al., 2020). It contains more than 600 000 unique sources detected in more than 12 000 observations that cover 3 % of the sky, 19 % of the sources are detected more than once. The detections have a median flux of 2.3×10^{-14} erg cm^{-2} s^{-1} and a median positional accuracy < 1.6 arcsec, and for about 23 % sources, the fluxes are below 1×10^{-14} erg cm^{-2} s^{-1},.

In addition, the XMMSL catalog provides observations detected during the slews of the satellite. It was updated multiple times by adding new slews as the mission progresses, the most recent version, XMMSL2, contains the data of 2114 slews executed between August 2001 and December 2014 that cover about 84 % of the sky. The detected sources are released in two different catalogs. The full catalog contains 72 352 detections, while up to 20 % of them might be spurious. Furthermore, 29 393 detections with a higher detection likelihood and that remain after the removal of suspicious sources, e.g. detections in the wings of the PSF of bright sources, are released as clean catalog. About 4 % of these detections are expected to be spurious. Since some of the slews overlap, several sources are detected multiple times, and thus, the detections of the clean catalog originate from 23 252 unique sources. The field of view of *XMM-Newton* and the slew speed of 90°/h results in a quite low exposure time with an average of about 7 s, which causes a detection limit of $\sim 5 \times 10^{-13}$ erg cm^{-2} s^{-1} for stellar sources (see Freund et al. (2018) (Sect. 3.1)). According to Saxton et al. (2008), the positional accuracy of the XMMSL2 sources is typically about 8 arcsec. Saxton et al. (2008) provide identifications of the XMMSL2 sources that are obtained by a crossmatch with different catalogs as the SIMBAD and NED database and the catalogs of galaxy clusters by Abell and Zwicky. Searching radii between 12 and 180 arcsec were applied depending on the positional accuracies and the sources extent of the matching catalog. As a result, identifications and classifications are provided for about 70 % of the XMMSL2 sources, however, the classification is unspecific, e.g. “X-ray”, for some sources. An improved identification of the stellar sources in the XMMSL2 catalog is presented by Freund et al. (2018) (see Sect. 3.1).

2.1.4 eROSITA

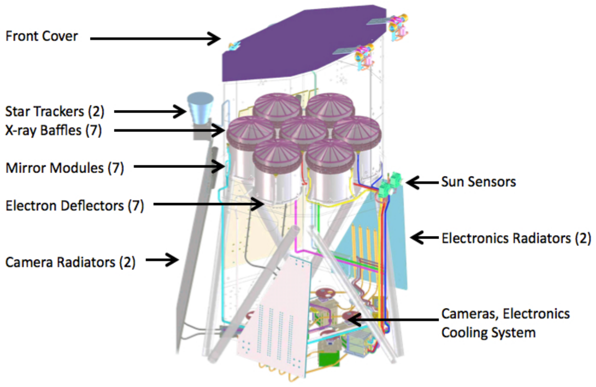


Figure 2.4: Schematic view of the *eROSITA* X-ray telescope. Adopted from Fig. 2.4.2 in Merloni et al. (2012)

This description mainly follows Merloni et al. (2012) and Predehl et al. (2021).

The structure of *eROSITA* is schematically shown in Fig. 2.4. In contrast to previous X-ray telescopes, *eROSITA* was not inserted into Earth orbit but instead in a halo orbit around the second Lagrangian point (L2) of the Earth-Sun system, i.e., a point 1.5 million km in the anti-sun direction were the gravitational and centrifugal forces of the Sun and the Earth balance each other.

The *eROSITA* instrument contains 7 identical co-aligned mirror assemblies with each consisting of 54 mirror shells. The advantage of the multi mirror concept is a 7-fold redundancy and a reduction of the focal length causing a reduced instrumental background and a more compact telescope. The mirror shells are made of electroformed nickel with reflective layers of gold. The telescopes achieve an on-axis angular resolution of about 16 arcsec. As typical for a Wolter type I telescope, the PSF rapidly degrades at higher off-axis angles leading to an angular resolution average over the field of view of about 26 arcsec. In the focal plane of each of the mirror assemblies, a CCD camera with 384 × 384 pixels covers the field of view with a diameter of 1°.03. The *eROSITA* CCDs are very similar to the pn camera on-board of *XMM-Newton* (see Sect. 2.3) but the *eROSITA* CCDs contain a frame store area that substantially reduces the number of events that are normally lost, when recorded during the read-out time. A massive copper shield protects the CCDs against particle radiation. For spectral calibration, every CCD module is equipped with a radioactive ^{55}Fe source and an Al/Ti target that provide spectral lines at 1.5, 4.5, and 5.9 keV. The cameras are protected by aluminum filters, which are deposited directly on the CCD for five cameras and for two cameras, it is located in a filter wheel. The effective area of the seven *eROSITA* telescopes is similar to the three *XMM-Newton* telescopes, but the so-called Grasp, which is the product of the field of view and the average effected area is about four times larger at 1 keV.

After its launch, *eROSITA* performed a calibration and Performance Verification (PV) phase testing many aspects of the instruments and the different observing modes. *SRG/eROSITA* can operate in pointing modes targeting a single sky position, in field scan mode, where a sky region is scanned in a rectangular grid pattern, and in survey mode, in which *eROSITA* scans great circles of the sky by continuous rotations of the spacecraft. The longest observation during the PV phase was dedicated to the *eROSITA* Final Equatorial Depth Survey (eFEDS). For eFEDS an area of about 140 deg^2 roughly located between 20° and 40° Galactic latitude, was observed in four rectangular field scans. The eFEDS field has a nearly uniform vignetting corrected exposure of approximately 1.2 ks, which is about 50 % deeper than expected after four years all-sky survey in the Ecliptic equator (Brunner et al., 2021). Optical identifications and classifications are provided by Salvato et al. (2021) (see Sect. 5.2) and methods specifically trained for stellar sources are described by Schneider et al. (2021) (see Sect. 5.1).

On December 12th, 2019 the first out of eight *eROSITA* all-sky surveys (eRASS) was started. Similar to *ROSAT*, *eROSITA* scans great circles over the Ecliptic poles during the survey. Due to the scanning rate of 90°/h and the angular velocity of the spacecraft around the Sun of about 1°/d, a source at the Ecliptic plane is observed six times in each of the eight all-sky surveys with the longest passage time being 40 s. One survey

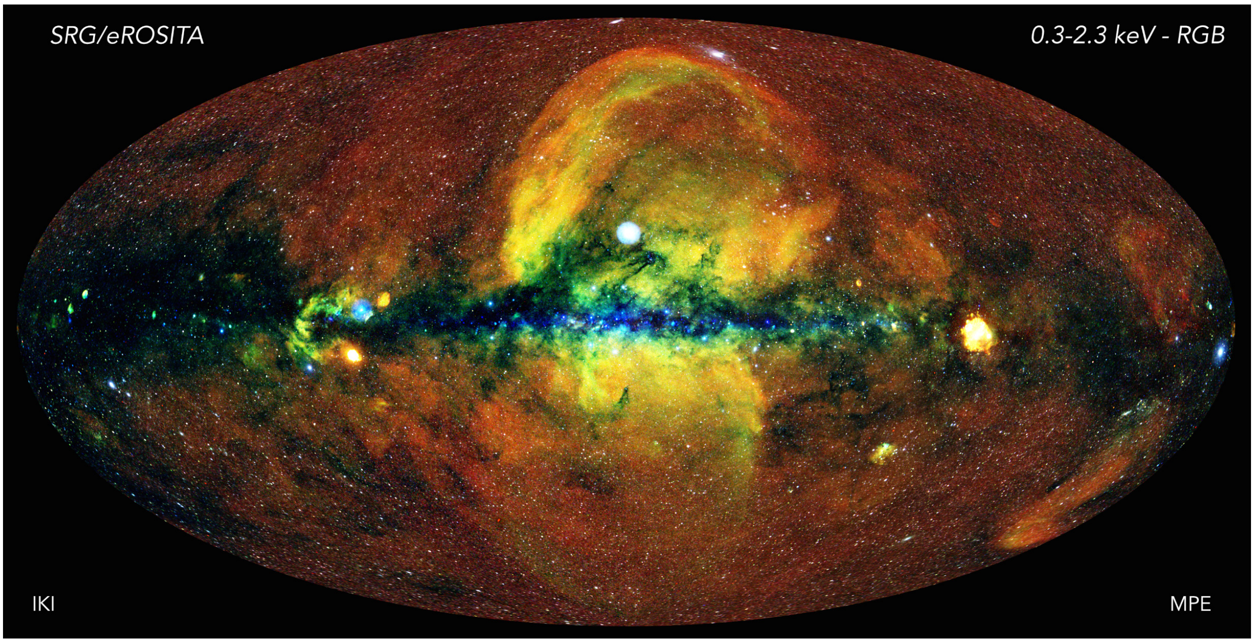


Figure 2.5: X-ray sky as obtained by the first *eROSITA* all-sky survey. Credit: Jeremy Sanders, Hermann Brunner and the eSASS team (MPE); Eugene Churazov, Marat Gilfanov (on behalf of IKI)

of the full sky is completed within half a year so that the eight surveys will be finished by the end of 2023. By then, the whole sky is expected to be observed with an average vignetting corrected exposure time of about 1.3 ks and an average sensitivity of $\sim 1.1\times10^{-14}$ erg cm^{-2} s^{-1} in the soft band (0.5-2 keV). Due to the scanning law, the exposure time and sensitivity are substantially higher near Ecliptic poles than in the plane. In Fig. 2.5 the image obtained from the first *eROSITA* all-sky survey is shown. After the all-sky surveys, 3.5 years of operation in pointed and scanning mode are planned.

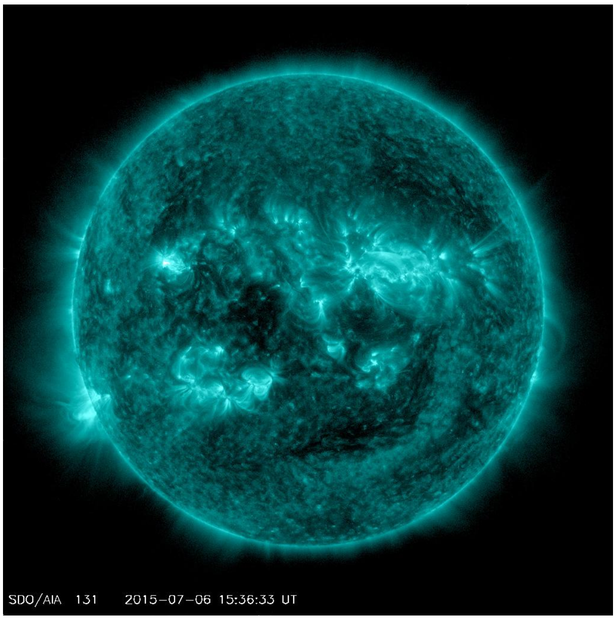


Figure 2.6: Image of the Sun at 131 Å (95 eV) from NASA's Solar Dynamics Observatory. Credit: NASA/GSFC/SDO

2.2 Properties of coronal X-ray sources

Reviews about X-ray astronomy of stellar coronae are provided by Güdel (2004) and Testa et al. (2015) that this description largely follows.

2.2.1 Production of coronal X-ray emission

First X-ray observations of the Sun during rocket flights and the *Skylab* mission in the early seventies revealed that the solar corona is highly structured (e.g. Vaiana et al. (1973)). In X-ray bright active regions, plasma is confined in coronal loops and heated to millions of Kelvin. The active regions tend to be located in two strips north and south of the solar equator and cover about 0.01-1 % of the solar surface. The plasma escapes the Sun as solar wind in coronal holes. Here, the temperature of the corona is much lower, and hence, these regions are darker in X-ray images. The structure of the corona continuously changes. In Fig. 2.6 an image of the corona of the Sun in the extreme ultraviolet (EUV) regime is shown, active regions and coronal holes are clearly visible.

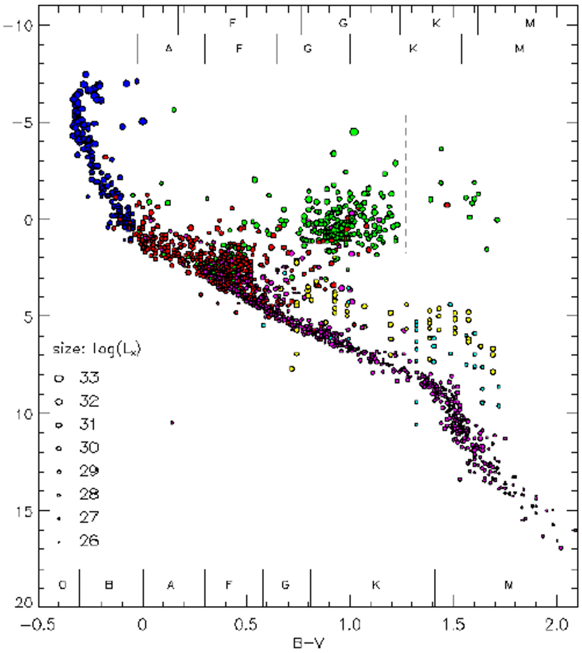


Figure 2.7: Color-magnitude diagram for about 2000 stellar X-ray sources adopted from different catalogs. The ranges of the spectral type for supergiants, giants and dwarfs are shown at the top and bottom of the figure and the size of the dots scales with the X-ray luminosity. Adopted from Fig. 2 in Güdel (2004)

While the Sun allows to study a corona at high spatial and spectral resolution, the dependence of the X-ray activity on stellar masses, ages, and rotation periods can be studied by observing stars. The *Einstein* mission and later *ROSAT* showed that X-ray emitters are found for nearly all types of stars (e.g. Schmitt et al., 1995; Schmitt, 1997; Huensch et al., 1998a,b). In Fig. 2.7 a color-magnitude diagram (CMD) of about 2000 stellar X-ray sources extracted by Güdel (2004) from different catalogs is shown (see Fig. 11 in Freund et al., submitted in Sect. 4.1 for an X-ray selected CMD of a coherent input sample). Main sequence X-ray sources from spectral type O down to the bottom of the main sequence are shown, and furthermore, giants and pre-main sequence stars are detected in X-rays. Early A-type stars without outer convection zone are generally thought to be X-ray dark (Schmitt et al., 1985), and X-ray emission detected for these objects is in most cases produced by an unresolved late-type companion. Stars of OB-type exhibit X-ray emission but it is produced by stellar winds and not in a corona (Pallavicini et al., 1981; Berghoefer et al., 1997). Even for stars of the same spectral type, the X-ray luminosities can differ by orders of magnitude. For example, the X-ray luminosity of the Sun is between 3×10^{26} and 5×10^{27} erg s^{-1} (Peres et al., 2000) but the solar-analog EK Dra at an age of about 70 Myr exhibits an X-ray luminosity of 1×10^{30} erg s^{-1} (Guedel et al., 1995).

The extremely high X-ray luminosities of some stars raise the question to what extent the stellar activity phenomena can be explained by the coronal features observed on the Sun. The minimum X-ray flux detected for stars agrees well with that of solar coronal holes (Schmitt, 1997), but for example, large polar spots were found on active stars (e.g. Berdyugina et al., 1998; Hussain et al., 2007), while spots and active regions on the Sun are located near the equator. Furthermore, the high-level of X-ray emission of some stars cannot be explained by a corona entirely covered with solar active regions, instead for instance, higher coronal plasma densities are observed.

The heating of the corona was initially tried to be explained by dissipation of acoustic waves (Schwarzschild, 1948). However, this model was inadequate to explain the inhomogeneity of the solar corona and the wide range of X-ray luminosities of stars with the same spectral type. Therefore, magnetic heating was proposed and Pevtsov et al. (2003) show that the logarithm of the X-ray luminosity of active stars and different solar coronal features is roughly proportional to the logarithm of magnetic flux over about 12 orders of magnitude.

The magnetic field of the Sun and other late-type stars is thought to be produced in a $\alpha\omega$-dynamo in the interior of the star (see e.g. Parker, 1955, 1979). A poloidal magnetic field is produced between the inner radiative core and the twisted motions of the outer convection zone (α-effect). The differential rotation then shears and amplifies the poloidal field to a toroidal field (ω-effect). However, the details of the coronal heating are still under debate, although the dissipation of Alfven waves (van Ballegooijen et al., 2011) or nanoflares may play a fundamental role (Parker, 1988). At the boundary to fully convective stars of late M-type, a transition of the dynamo is expected but no change of the X-ray activity is observed (Fleming et al., 1995).

In addition to the X-ray and magnetic activity, there are also chromospheric activity indicators that are, for example, based on the emission in the Ca II HK lines. Many of the properties of X-ray activity were actually first studied with chromospheric activity indicators. A review about stellar chromospheric activity is provided by Hall (2008). For late-type stars and various solar coronal features, the X-ray and chromospheric line fluxes follow a relationship over 2-4 orders of magnitudes (Schrijver & Zwaan, 2008). Together with the results of Pevtsov et al. (2003) this suggests that a common heating mechanism is present for all solar and stellar coronal features.

2.2.2 Activity-rotation-age relation

Originally, with Ca II HK data, Skumanich (1972) showed that the most important parameter that defines the stellar activity is the stellar rotation. Later, this relation was studied in X-rays (Pallavicini et al., 1981; Walter & Bowyer, 1981; Pizzolato et al., 2003; Wright et al., 2011, 2018). As shown, e.g., by Wright et al. (2011) in Fig. 2.8, the X-ray to bolometric luminosity ratio linearly increases with decreasing rotation period in the double logarithmic plot for slow rotators. For stars with a short rotation period, the luminosity ratio saturates and does not further increase for faster rotators (Vilhu & Rucinski, 1983). Due to the saturation limit of the X-ray luminosity, the optical brightness of counterparts to flux limited X-ray surveys is also restricted (see Fig. 1 in Freund et al., 2018, (Sect. 3.1)). There is some dependence of the stellar mass on the saturation threshold because the X-ray activity depends not only on the rotation period P_{rot} but also on the convective turnover time τ_{conv} (Noyes et al., 1984a). With the Rossby number $R_O = P_{\mathrm{rot}}/\tau_{\mathrm{conv}}$, the X-ray to bolometric luminosity ratio can be parameterized by the ansatz

$$R_X = \begin{cases} R_{X\mathrm{sat}} & : \ R_O < R_{O\mathrm{sat}} \\ C R_O{}^{\beta} & : \ R_O > R_{O\mathrm{sat}} \end{cases}, \tag{2.1}$$

where Wright et al. (2011) estimate for the saturation limit $R_{X\mathrm{sat}} \approx -3.13$. This value seems to be independent of the stellar color in their sample. (Similar results for stars of equal age are presented by Freund et al. (2020) in Sect. 6.1.)

The physical causes of the saturation are still under debate. Possible explanations of the saturation include the saturation of the dynamo efficiency (Gilman, 1983), the complete coverage of the stellar surface with active regions (Vilhu, 1984), or a change of the underlying dynamo mechanism (Barnes, 2003). For some extremely fast rotators ($R_O \lesssim 0.01$), the X-ray activity seems to decrease again (Randich et al., 1996). This so-called supersaturation effect is not observed for M-type dwarfs but only for earlier type stars. The supersaturation might be caused by stripping of coronal loops due to the high centrifugal forces in rapid rotators (Jardine, 2004) or by strong polar updrafts in the convection zone resulting in a field-free equatorial region (Stępień et al., 2001).

A strong relation of the stellar activity on age is also observed (Skumanich, 1972), with young stars exhibiting large X-ray luminosities that monotonically decay with age. However, this decrease is a consequence of the increasing

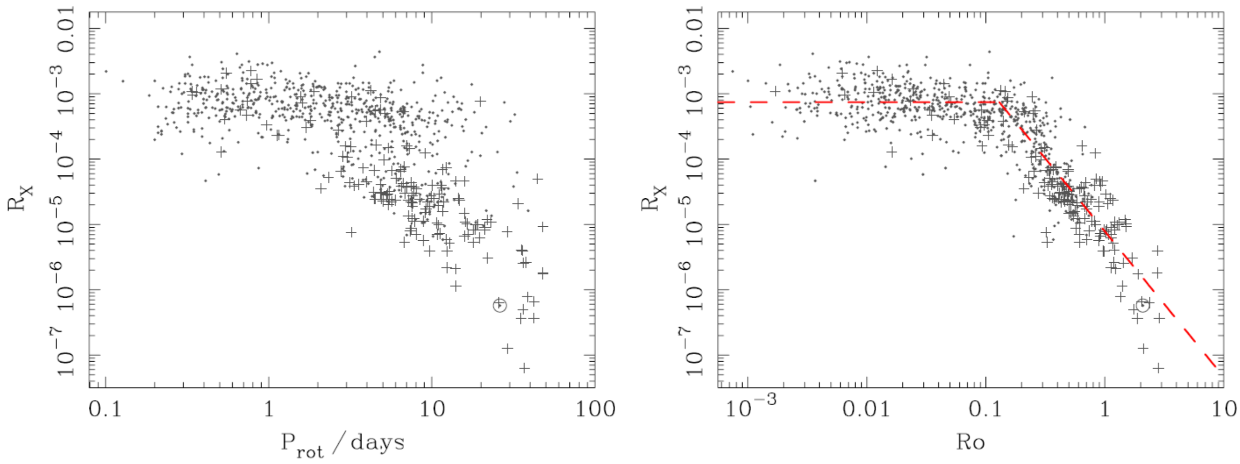


Figure 2.8: Relationship between the X-ray to bolometric luminosity ratios R_X and the rotation periods (left) and Rossby numbers (right). Adopted from Fig. 2 in Wright et al. (2011)

rotation period with age (Hempelmann et al., 1995). The rotation period of zero-age main sequence stars largely vary, which might be caused by a different star-disk coupling or disk-dispersal histories. Nevertheless, the rotation is generally fast enough for saturation of the X-ray emission (e.g. Patten & Simon, 1996). The strong magnetic fields produce winds that cause magnetic breaking, and hence, slowing the rotation rate. This negative feedback loop reduces the X-ray activity with increasing stellar age and the initial spread of the rotation periods converge. The spin-down rate is highest for the earliest-type stars. Consequently, F-type stars are (if at all) saturated for a very short time, G- and K-type dwarfs become unsaturated after roughly 50 and 100 Myr, and M-type stars remain saturated for several hundred Myr (see Güdel, 2004, and references therein). (The X-ray activity of the 640 Myr old Hyades is discussed in Freund et al. (2018) (Sect. 6.1).)

When late-type stars with masses $\lesssim 1.5\ M_\odot$ evolve towards the giant branch, their X-ray luminosity further decreases because they developed to slow rotators during their time on the main sequence and the rotation period further decreases as the radius increases (Pizzolato et al., 2000). However, O-, B-, and A-type stars do not spin down on the main sequence and when they evolve to giants, convection sets in and X-ray luminosities of up to 10^{31} erg s^{-1} are reached (Maggio et al., 1990). A magnetized wind again leads to a braking of the stellar rotation but at the same time, the convection zone deepens. Therefore, no unique rotation-activity relation exists for giants (Gondoin, 1999). At spectral type K3 between luminosity classes II-IV, the so-called dividing line separates X-ray bright giants with a hot corona to the left and X-ray faint stars that emit massive cool winds to the right (Linsky & Haisch, 1979; Huensch et al., 1996, see also the dashed line in Fig. 2.7). However, bright giants and supergiants located to the right of the dividing line can exhibit both, a hot corona with X-ray emission and cool winds, they are called hybrid stars (Reimers et al., 1996).

Due to tidal interaction, close binaries can maintain their high rotation period during their time on the main sequence and giant evolution, and hence, they are strong X-ray emitters. A typical example of these active binaries are RS CVn-type systems that contain a G- and K-type giant or subgiant and a late-type subgiant or main sequence companion. X-ray luminosities between 10^{29} and a few times 10^{31} erg s^{-1} are found for RS CVn system (Walter & Bowyer, 1981; Dempsey et al., 1993), but also similar systems like BY Dra- or Algol-type binaries exhibit high X-ray activity. Contact binaries of type W UMa have even higher rotation periods but their X-ray to bolometric luminosity ratios are reduced (McGale et al., 1996).

2.2.3 Long and short term variability

The X-ray luminosity of stars can increase by orders of magnitude on short time scales due to flares. Flares are thought to be produced by the reconnection of neighboring antiparallel magnetic field lines at large coronal heights. Thereby, electrons are accelerated and when they hit the chromosphere, the plasma is heated to MK. The hot plasma evaporates into the corona, where it cools through conduction and radiation. The flare process is observed in multiple parts of the electro magnetic spectrum; from radio gyrosynchrotron radiation through optical radiation to hard and soft X-ray emission.

Originally for solar observations, flares are categorized in two types; in compact flares, individual coronal loops of modest height and high densities brighten up within minutes, while long-duration flares, also named two-ribbon flares, have a more complex structure, larger sizes, lower densities, and last for hours (Pallavicini et al., 1977). Generally, all coronal active stars show flares, and a large variety of stellar flares is observed with durations ranging from a few minutes to more than ten hours and temperatures of up to 100 MK (see Güdel, 2004, and references therein). Most flares show a sharp increase of the luminosity and a long exponential decay but also flares with reheating events and secondary peaks are reported (e.g. Katsova et al., 1999; Güdel et al., 2004). Highly active and very late-type stars exhibit stronger and more frequent flares. A particularly strong flare is reported by Stelzer et al. (2006) who observed an increase of the X-ray luminosity by a factor of 200 - 300 compared to the quiescent emission and a difference in the optical brightness of $\Delta V = 6$ mag on the dwarf LP 412-31 of spectral type M8. Stars detected during strong flares in X-ray surveys may apparently be located above the saturation limit especially when the exposure time is short (cf. Freund et al., 2018, in Sect. 3.1).

However, such large flares are very rare and the number of flares with a specific energy is described by a power-law ($dN/dE \propto E^{-\alpha}$; see e.g. Lin et al. (1984) for the solar case). Due to the increasing number of small flares, there is often no true quiescent emission but the light curves consist of many small and unresolved flares. For a value of $\alpha > 2$, the total energy release is dominated by small flaring events. In this case the coronal heating could be described by so-called “nanoflares”. This idea is supported by the fact that stellar quiescent luminosity and the flare rate show a nearly linear relation (Audard et al., 2000). However, the value of α is still under debate (Audard et al., 2000; Kashyap et al., 2002).

On the timescale of years, the X-ray luminosities of stars change due to activity cycles. The luminosity of the Sun in soft X-rays varies by more than one order of magnitude (Peres et al., 2000) with a period of 11 years, although neither the amplitude nor the cycle length are constant. During the solar maximum the stripes of active regions broaden and in the minimum, they almost completely disappear. Extensive observations of chromospheric activity indicators through the Mount Wilson program of Ca II HK emission found activity cycles for many late-type stars, while slow rotating and low active stars generally have smoother cycles (Baliunas et al., 1995). The detection of stellar activity cycle in X-rays suffers from the high intrinsic variability caused by flares and the lack of continuous X-ray monitoring. Robrade et al. (2012) report on the detection of activity cycles for the weakly to moderately active binaries 61 Cyg A/B and α Cen A/B of G and K type during observations over almost 10 years with *XMM-Newton.* They observe variations in X-ray brightness with a factor of 3 to 10 over periods of about 7 to 15 yr for the different stars. There are indications that the cycle period increases with the rotation period as $P_{\rm cyc} \propto P_{\rm rot}^{1.25}$ (Noyes et al., 1984b). (In Freund et al. (2020) (Sect. 6.1) the variability of the 640 Myr old Hyades members is analyzed.)

2.3 Gaia mission

The most important resource for optical counterparts in the context of this thesis is *Gaia*, a cornerstone mission of ESA, that operates at the second Lagrange point (L2) of the Earth-Sun system. It provides parallaxes and magnitudes in three different bands for the stellar X-ray sources and information about the object classification and stellar parameters are planned for the future. Extension to the nominal five-year mission lifetime of *Gaia* has been approved at least until the end of 2022.

2.3.1 Spacecraft and instruments

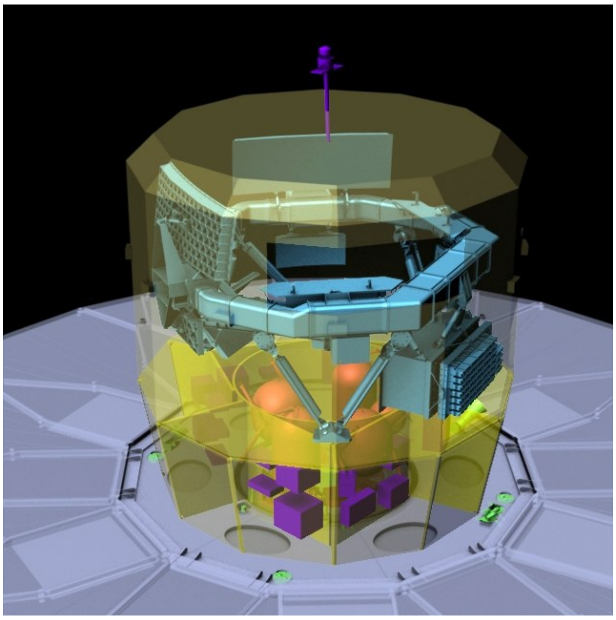


Figure 2.9: Schematic view of the *Gaia* spacecraft and its modules. Credit: ESA/AOES Medialab

Gaia's instruments and mission is described in detail by Gaia Collaboration et al. (2016b) and I summarize the most important aspects in the following. An artist's impression of the components of *Gaia* is shown in Fig. 2.9.

Gaia contains two identical telescopes with apertures of 1.45 × 0.50 m. The fields of view of the telescope are separated by the basic angle of $\Gamma = 106.5^\circ$ and are combined on a shared focal plane. The light of the telescopes is reflected six times to accommodate the focal length of 35 m, the fourth mirror combines the beams of the two telescopes. The spacecraft rotates with a speed of 60 arcsec s^{-1} around the spin axis that is directed at an angle of 45° to the Sun and precesses with a period of 63 days. The combination of both rotations in addition to the movement around the Sun leads to a coverage of the full sky with, on average, 70 observations of each object during the norminal five year mission. Due to the rotation around the spin axis, one-dimensional positions in the along-scan direction are derived from the observation times. The precession of the spin axis results in an intersection of the scanning planes at large angles allowing the determination of two-dimensional coordinates from the one-dimensional measurements. Thus, highly accurate angular separations between stars in a given field of view are obtained. However, the parallax factors of these sources are nearly identical, and therefore, only relative parallaxes can be measured. Due to the two fields of view separated by a large basic angle, also separations of objects in the different fields of view are measured. The parallax factors of these sources substantially differ, and hence, absolute parallaxes are obtained.

The focal plane assembly consists of 106 CCDs with a total of almost one billion pixels, each covering 58.9 × 176.8 mas on the sky. The CCDs are dedicated to source detection, astronomy, low-resolution spectro-photometry, and spectroscopy. There are three different types of CCDs; while the general CCDs are optimized for detecting photons in a broad bandpass (330 – 1050 nm), the blue and red CCDs differ for example in their anti-reflection coating to optimized the detection in the wavelength range of 330 – 680 nm and 640 – 1050 nm, respectively. The CCDs are arranged in 7 rows and 17 strips, and each object passes every strip within approximately two minutes due to the rotation of the spacecraft.

Objects in the field of the view of *Gaia*'s telescopes first pass the Sky Mapper (SM) that consists of two strips with in total 14 broad band CCDs. Here, the objects are detected, with the first and second strip exclusively record the objects observed by the preceding and following telescope, respectively. In contrast to the subsequent fields, the SM CCDs are read out in full-frame mode to define observational windows around objects of interest. Only these windows are read in the following CCDs to reduce the generated data rate. Then, the 62 Astrometric Field (AF) CCDs provide astrometric measurements and white-light photometry in the G band.

Measurements of the spectral energy distribution (SED) are obtained by two fused-silica prisms called blue (BP) and red photometer (RP) that are mounted directly in front of the focal plane. The dispersed light illuminates two strips of 7 blue and 7 red type CCDs. The spectral dispersion varies for BP from 3 to 27 nm pixel^{-1} and for RP from 7 to 15 nm pixel^{-1} for the wavelength ranges of the blue and red type CCDs. Since the photometric instruments uses the same telescopes, the same focal plane, and the same sky mapper, SEDs are obtained for the same objects and transits as observed by the astrometric instrument. From the SEDs, classifications and parameters as interstellar reddening, metallicities, surface gravities, and effective temperatures are derived. Magnitudes in the BP and RP band are estimated from the integrated light of the spectra.

Furthermore, the radial-velocity spectrometer (RVS) obtains spectra with a dispersion of 0.0245 nm pixel^{-1} for the brightest *Gaia* sources. As the photometers, the RVS is mounted in front of the focal plane and contains a blazed-transmission grating plate and fused-silica prismatic lenses. A multilayer-interference bandpass filter restricts the wavelength range to 845 – 872 nm, which covers the Ca II triplet that provides, among other things, information about the chromospheric activity. The light dispersed by the RVS illuminates 12 red type CCDs, but in contrast to the other focal plane areas, they are not arranged in seven but in four rows and three strips. Hence, 57 % of the transits detected by the astrometric and photometric instruments are observed by the RVS. For objects brighter than about 16th magnitude in the RVS bandpass, radial velocities and other parameters are estimated from the RVS spectra.

In addition, *Gaia* is equipped with a basic angle monitor to measure variations of the basic angle between the two fields of view at μas level and wave-front sensors to monitor the optical performance of the telescopes. Based on science data and data from the wave-front sensors, the secondary mirrors can be adjusted to improve the performance of the telescopes.

2.3.2 Catalogs

Three *Gaia* data releases with large improvements have been published so far, and all of them are used in the articles presented in this thesis. I briefly summarize the contents of the releases and the expectations for the final *Gaia* catalog.

***Gaia* DR1**

In September 2016, the first *Gaia* data release (*Gaia* DR1; Gaia Collaboration et al., 2016a) was published, and it is used in Freund et al. (2018) (see Sect. 3.1). It is based on the data taken in the first 14 month of operation and contains positions with an accuracy of about 10 mas (Lindegren et al., 2016) and magnitudes in the broad G band (Evans et al., 2017) for more than 1.1 billion sources. For a subsample of more than 2 million stars in common between Gaia DR1, Hipparcos and Tycho2, called Tycho-Gaia astrometric solution (TGAS), positions with higher accuracy (see Table 2.2), proper motions, and parallaxes are provided (Lindegren et al., 2016). Furthermore, G band light curves of more than 3000 RR-Lyrae and Cepheids located near the south ecliptic pole are provided.

Due to the short observing time and a simplified data processing, many limitations are known for *Gaia* DR1 and it is incomplete for various reasons. The short time period taken into account causes non-astronomical artifact where the source density is substantially reduced in some sky regions because of a poor scan law coverage. Furthermore, many stars brighter than $G = 7$ mag (or sources close to bright stars), objects with high proper motions, and extremely blue and red sources are missing in *Gaia* DR1. In dense regions with more than 400 000 sources per square degree and for binaries with separations of less than ~ 4 arcsec, *Gaia* DR1 contains only the brighter objects. Positions, proper motions, and parallaxes might have larger errors because the orbital motion in binaries is ignored and a parallax zero point of ±0.1 mas was found (Lindegren et al., 2016). These limitations are particularly relevant for the identification of X-ray sources because stellar X-ray emitters are often optically bright, close to the Sun (and therefore have large proper motions), and frequently occur in binary systems.

Gaia DR2

The second *Gaia* data release (*Gaia* DR2; Gaia Collaboration et al., 2018a), published in April 2018, is relevant in the context of Freund et al. (2020) (see Section 6.1). It is constructed from observations taken within 22 months of observation with an improved data processing compared to *Gaia* DR1. *Gaia* DR2 contains positions and G band magnitudes for almost 1.7 billion sources. Five-parameter astrometry with parallaxes and proper motions independent from Tycho2 and Hipparcos data is provided for more than 1.3 billion brighter than $G = 21$ mag (Lindegren et al., 2018). Furthermore, magnitudes in the BP and RP bands are given for nearly 1.4 billion *Gaia* DR2 sources (Riello et al., 2018; Evans et al., 2018). For more than 7 million sources approximately brighter than $G = 13$ mag and with effective temperatures between 3550 and 6900 K, radial velocities are provided (Katz et al., 2019). *Gaia* DR2 also contains effective temperatures in the range of 3 000 to 10 000 K for about 160 million stars brighter than $G = 17$ mag and extinction, reddening, radius, and luminosity values are given for about half of these sources (Andrae et al., 2018). In addition, *Gaia* DR2 provides photometric time series for 550 000 variable sources categorized into 9 variable types (Holl et al., 2018), and astrometry and photometry are given for 14 000 solar system objects (Gaia Collaboration et al., 2018b).

Many of the limitations known for *Gaia* DR1, such as incompletenesses of or near bright stars and reduced source densities in some sky regions due to the scanning law, are considerably reduced but still present in *Gaia* DR2. About 17 % of the sources with proper motions > 0.6 arcsec yr^{-1} are missing, and in regions with a few hundred thousand objects per square degree, the magnitude limit of *Gaia* DR2 can be as bright as $G = 18$ mag. The angular resolution of *Gaia* DR2 is about 0.4 arcsec but companions in close pairs begin to be missing at separations below about 2 arcsec. A parallax zero point of -30 μas is determined, and furthermore, some sources have erroneous highly significant positive or negative parallaxes caused by source confusion. They can be filtered out based on the goodness of the astrometric fit (Lindegren et al., 2018).

Gaia DR3

The third *Gaia* data release, based on data collected during the first 34 months of the mission, was split in two releases. The early data release (*Gaia* EDR3; Gaia Collaboration et al., 2021) released in December 2020 contains the same astrometric and photometric parameters as *Gaia* DR2 but with an improved accuracy and for a larger number of sources. It is used by Freund et al., submitted and Schneider et al. (2021), see Sects. 4.1 and 5.2, respectively. *Gaia* (E)DR3 contains 1.8 billion sources, and for more than 1.4 billion objects, parallaxes and proper motion are provided (Lindegren et al., 2021) and 1.5 billion sources contain photometry in the BP and RP band (Riello et al., 2021). The full *Gaia* DR3 is expected in the second quarter of 2022 and will complement the early data release with BP/RP and RVS spectra for 100 and 1 million objects, respectively. From these spectra, radial velocities, object classifications, and astrophysical parameters are derived for about 33 million, 1 billion, and 500 million sources, respectively, and a few hundred thousand sources will be identified as non-single stars. Furthermore, *Gaia* DR3 will contain about 30 million variable source classifications and 150 000 solar system objects.

The data processing of *Gaia* EDR3 is generally the same as for *Gaia* DR2, but a color-dependent calibration of the point-spread function have been applied for the first time to almost 585 million *Gaia* EDR3 sources with a well-determined color in *Gaia* DR2. For further 882 million predominantly faint sources, a pseudo-color was estimated as the 6th astrometric parameter. The accuracy of the 5-parameter solution is generally higher (Lindegren et al., 2021). Compared to *Gaia* DR2, the completeness at the faint end and in close binaries is improved in *Gaia* EDR3, and furthermore, the number of spurious highly significant parallaxes and proper motion is reduced. A detailed description of

Table 2.1: Contents of the different *Gaia* data releases

	Gaia DR1	*Gaia* DR2	*Gaia* DR3
Total number $[10^6]$	1 143	1 693	1 812
5-parameter $[10^6]$	2.054	1 332	1 468[a]
G magnitudes $[10^6]$	1 143	1 693	1 806
BP magnitudes $[10^6]$	-	1 382	1 542
RP magnitudes $[10^6]$	-	1 385	1 555
Radial velocities $[10^6]$	-	7.225	33[b]
BP/RP spectra $[10^6]$	-	-	> 100[b]
RVS spectra $[10^6]$	-	-	1[b]
Astrophysical parameters $[10^6]$	-	161	500[b]
Variable classifications	3 194	551×10^3	13×10^6[b]
Solar system objects $[10^3]$	-	14	150[b]

[a] 5- and 6- parameter solutions combined

[b] Expected value for *Gaia* DR3

Table 2.2: Typical uncertainties for the different *Gaia* data releases

	Gaia DR1	*Gaia* DR2	*Gaia* DR3	end-of-mission[a]
Position at $G = 15$ mag [μas]	300[b]	40	25[c]	20
Parallax at $G = 15$ mag [μas]	300[b]	40	35[c]	27
Proper motion at $G = 15$ mag [μas/yr]	1000[b]	70	35[c]	14
Systematic astrometric errors [μas]	300	< 100	< 50	-
G band at $G = 17$ mag [mmag]	< 30	2	1	0.24
BP band at $G = 17$ mag [mmag]	-	10	12	1.5
RP band at $G = 17$ mag [mmag]	-	10	6	1.1

[a] Estimation of the performance at the end of the nominal five-year mission for a sun-like star as given by Gaia Collaboration et al. (2016b)

[b] For TGAS sources

[c] Average between 5- and 6- parameter solution

the validation of the *Gaia* EDR3 catalog and the found shortcomings is provided by Fabricius et al. (2021).

Comparison

In Table 2.1 I compare the number of sources for which the different parameters are provided in the various *Gaia* data releases. The number of parameters and the number of sources for which the parameters are available constantly increased with every new data release. In Table 2.2 I compare the typical uncertainties of the parameters in the different Gaia releases as given by Gaia Collaboration et al. (2016a), Gaia Collaboration et al. (2018a), Gaia Collaboration et al. (2021), and Gaia Collaboration et al. (2016b). The accuracy of each parameter increased with every data release and is expected to further improve to the end of the mission. Only the accuracy of the magnitude measurement in the BP band seems to decrease from *Gaia* DR2 to DR3 but this is only apparent because the uncertainties in the BP and RP bands are not distinguished in DR2. Note that the end-of-mission values stated in Table 2.2 are the uncertainties expected for the nominal 5-year mission and that even higher accuracies will be obtained as the mission is extended. Naturally, the increase is the largest at the beginning of the mission and progressively more time is needed to obtain the same improvement.

2.4 Bayesian statistics and the matching problem

Before the X-ray properties of stars can be investigated, the stellar content of the X-ray catalogs needs to be identified. In Sect. 2.4.1 I outline identification procedures previously applied to find stellar counterparts to X-ray sources. Most of these methods only consider the geometric distances between X-ray sources and selected optical counterparts. Nowadays, approaches of Bayesian statistics are popular, especially, when additional properties, e.g., the counterpart magnitude, needs to be considered. In the Bayesian framework, prior knowledge is updated as more data or evidence become available. In Sect. 2.4.2 I summarize the most import aspects of Bayesian statistics to this work. In Sect. 2.4.3 I briefly summarize the Bayesian matching procedure, more details are provided below by Freund et al., submitted (see Sect. 4.1) and Schneider et al. (2021) (see Sect. 5.1). A similar approach is described by Budavári & Szalay (2008) and applied by Salvato et al. (2018) and Salvato et al. (2021) (see Sect. 5.2) to identify sources of the XMMSL, RASS, and eFEDS catalogs with the crossmatching tool NWAY.

2.4.1 Previous identification procedures

A very extensive identification of X-ray sources was performed for the *Einstein Observatory* Extended Medium-Sensitive Survey (EMSS; Gioia et al., 1990). The EMSS contains 835 X-ray sources serendipitously detected with *Einstein*'s IPC in 1435 pointings that are centered outside the Galactic plane ($|b| > 20°$). To avoid a bias by the selection of pointing targets, objects within 5 arcmin to the target object were excluded as well as fields centered on groups of targets, e.g., galaxy clusters and stellar associations. Finally, the EMSS covers an area of 778 square degrees and has a limiting sensitivity between about 5×10^{-14} and 3×10^{-12} erg cm^{-2} s^{-1}. For more than 96 % of the EMSS sources, identifications and classifications are provided (Stocke et al., 1991). To find the correct identification, spectra of counterparts within the 90 % confident error circle (typically between 35 to 70 arcsec) of the EMSS positions were taken until an optical counterpart with a reasonable source type and X-ray to optical flux ratio was found. For stellar sources, also the Ca II H&K emission and the projected stellar rotation speed were analyzed. According to Stocke et al. (1991), the obtained identifications are reliable to 97.5 %.

The identification of the EMSS sources was very time consuming and it took seven years to obtain the necessary spectra. Clearly, such an identification procedure is not feasible for larger surveys with tens or hundreds of thousands of X-ray sources like the *ROSAT* or *eROSITA* all-sky survey. Therefore, cross-correlations of samples with specific sources of interest were performed. For example, Dempsey et al. (1993) searched for an excess of X-ray photons in the RASS in the vicinity of 136 known RS CVn systems (a method previously adopted by Fleming et al. (1993) to identify X-ray emission of late M dwarfs). Similarly, X-ray emission from nearby late type giants (Huensch et al., 1996) and dwarfs of spectral type A to G (Schmitt, 1997) and K and M (Schmitt et al., 1995) were investigated including also data from *ROSAT* pointings. Larger input samples were positionally crossmatched with RASS sources detected independently of the optical position. Berghoefer et al. (1996, 1997) identified 237 of 1822 bright OB stars with RASS sources within a matching radius of 150 arcsec. Matching circles of 90 arcsec were applied by Huensch et al. (1998a) to identify 450 out of 3829 bright giants and supergiants and by Huensch et al. (1998b) who find 980 out of 3054 A, F, G, and K dwarfs and subgiants. A volume limited sample is presented by Schmitt & Liefke (2004) who find an identification for 1333 of 3231 stars within 25 pc adopting a search radius of 120 arcsec for the RASS and further data from pointed *ROSAT* observations. Volume-limited samples are more appropriate to analyze the average properties of stellar X-ray sources, while flux-limited samples contain in particular the high luminosity tail of the distribution because sources with a high X-ray luminosity can be detected to larger distances. A large sample of stellar RASS sources is provided by Guillout et al. (1999), who find a RASS source within 30 arcsec for 13 875 of about 1 million Tycho stars.

Due to the relatively small number of potential counterparts for any given X-ray source, the probability of finding a target source by chance in the search radius is small. Since also the fraction of sources of interest detected in X-rays is large, the reliability of the obtained X-ray identifications is high, even if a large matching circle is applied. However, the samples are biased by the preselection of input sources, and furthermore, the reliability decreases when the number of potential counterparts increases. For example, Guillout et al. (1999) expect about 7 % of their RASS-Tycho sample to be spuriously identified, although they adopted a comparatively small matching radius of 30 arcsec.

Rutledge et al. (2000) and Haakonsen & Rutledge (2009) additionally considered the counterpart's magnitudes in the B and J band to find associations to the brightest RASS sources in the USNO-A2 and 2MASS catalog, respectively. Specifically, they estimated for each counterpart a figure of merit for finding an association at the matching distance adopting only brighter sources as background objects. This implies that brighter counterparts are more likely identifications. Next, they compared the number of counterparts with this figure of merit for true and randomized RASS sources and estimated the probability that the counterpart is the correct identification. They provide samples of different identification quality, and the largest samples contain 11 301 USNO-A2 and 10 286 2MASS counterparts to 18 811 RASS sources, but the contamination of spurious associations is expected to be about 18 % and 15 %, respectively. Rutledge et al. (2000) and Haakonsen & Rutledge (2009) do not provide source classifications other than from the SIMBAD database. A Bayesian method that considers the color and magnitude of the counterpart in addition to the geometric match is applied by Salvato et al. (2018) to identify RASS and XMMSL2 sources with the cross-matching tool NWAY. They identify more than 63 000 and 13 000 RASS and XMMSL2 sources, respectively, and claim a reliability of more than 94 and 97 %. The method of Salvato et al. (2018) is trained for all point-like sources but with a special interest in AGN, and hence, they exclude sources within the Galactic plane ($|b| < 15^\circ$) and near the Magellanic Clouds. A detailed comparison of the RASS identifications from Salvato et al. (2018) with results of an identification procedure specialized for stars is provided by Freund et al., submitted, (Sect. 4.1).

2.4.2 Introduction to Bayesian statistics

A framework to consider additional properties of the match is the Bayesian statistics that was first described by Bayes & Price (1763) and further developed by LaPlace (1814). In contrast to the frequentist interpretation, probabilities are interpreted as a quantification of the degree of belief and are assigned to hypotheses and model parameters in the Bayesian framework. A detailed description of Bayesian statistics is provided for example by Loredo (1992) and Gelman et al. (2013).

In Bayes' theorem, the conditional probability of a hypothesis H is based on prior information and the obtained data. The prior probability $P(H)$ describes the probability of the hypothesis before the data are taken. Then, the compatibility of the (new) data D with the hypothesis H is indicated by the likelihood $P(D|H)$ of observing the data under the assumption that the hypothesis is true. The posterior probability of a hypothesis is estimated by Bayes' rule through

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}, \tag{2.2}$$

where the normalization factor $P(D)$ is called marginal likelihood.

When there are two competing hypotheses H_1 and H_0, the Bayes factor quantifies the support of one hypothesis over the other based on the obtained data D. It is estimated through

$$B = \frac{P(D|H_1)}{P(D|H_0)}, \tag{2.3}$$

where $P(D|H_1)$ and $P(D|H_0)$ describe the support of the data to the hypotheses. When the data favor H_1, the Bayes factor is larger than 1 and a value larger than 10 is often interpreted as strong evidence for the hypothesis H_1. However, the particular interpretation depends on the specific application and also involves the prior probability.

2.4.3 Applying Bayesian statistics to the matching problem

When we want to identify XMMSL, RASS, or eFEDS sources with their Gaia counterpart, we deal with the problem of finding the correct association to N_X X-ray sources in another catalog with N_O potential counterpart. For each of the N_X sources, I investigated $N_O + 1$ hypotheses, namely, $H_{ij},\ j = 1...N_O$ that the i^{th} X-ray source is associated to the j^{th} counterpart and H_{i0} that the i^{th} X-ray source has no identification in the counterpart catalog. Following the Bayesian framework, I first estimated the prior probability before any data were obtained. The probability of selecting the correct association by chance is the inverse of the number of counterparts if every X-ray source has an identification. However, when the counterpart catalog is incomplete or contains only a specific source type, the catalog fraction CF of X-ray sources with identification has to be taken into account and the prior probability is given by

$$P(H_{ij}) = \frac{CF}{N_O} = \frac{CF}{\eta\Omega}, \tag{2.4}$$

where η and Ω are the counterpart density and the solid angle of the full sky. On the other hand, the prior probability that the X-ray source is not associated with any of the counterparts is estimated through

$$P(H_{i0}) = 1 - CF. \tag{2.5}$$

Next, the positions of the X-ray sources and counterparts are considered. When the positional errors are Gaussian distributed with uncertainties of σ_X and σ_O for the X-ray sources and counterparts, respectively, the likelihood that the two sources are associated to the same physical object is estimated through

$$P(D_i|H_{ij}) = \frac{1}{2\pi(\sigma_X^2 + \sigma_O^2)}\, e^{-\frac{r_{ij}^2}{2(\sigma_X^2+\sigma_O^2)}}, \tag{2.6}$$

where r_{ij} is the angular separation between the X-ray source and the counterpart. Contrary, the likelihood that the X-ray source is not among the sample of potential counterparts and so a source is randomly within r_{ij} is given by

$$P(D_i|H_{i0}) = \frac{1}{4\pi}. \tag{2.7}$$

Applying Bayes' rule, the posterior probability that the i^{th} X-ray source is associated to the j^{th} counterpart is derived by the prior probability and the likelihood of the data through

$$P(H_{ij}|D_i) = \frac{P(D_i|H_{ij}) \cdot P(H_{ij})}{\sum_{k=0}^{N_O} P(D_i|H_{ik}) \cdot P(H_{ik})}. \tag{2.8}$$

To illustrate the influence of the parameters on the derived posterior matching probability, I show in Fig. 2.10 the matching probabilities versus the angular separation for different positional accuracies and counterpart densities. In all cases, I adopted a catalog fraction of 25 % and a negligible positional uncertainty of the counterparts, which is typical when matching X-ray sources and optical counterparts. Sources with a high positional accuracy and low counterpart densities, e.g., *eROSITA* sources near the Galactic pole, have very high matching probabilities for small angular separations and the transition to low probabilities at large separations is quite sharp. Therefore, true identifications and spurious associations are well separated. The probability of a match to be true stays high at large separations for sources with a low positional accuracy, e.g., RASS sources. However, for these sources, the association might be random even if the X-ray source and the counterpart have the same position. This is also true for sources with a high positional accuracy but a large counterpart density as for sources located in the Galactic plane. Low accurate sources with a high counterpart density always obtain small matching probabilities, and hence, such sources cannot be identified only by their positional match.

The quality of the identification can be improved when additional properties of the match, e.g., X-ray to optical flux, color, or counterpart distance, are considered by a Bayes factor. The Bayes factor compares the likelihoods of obtaining the properties for the cases that both sources

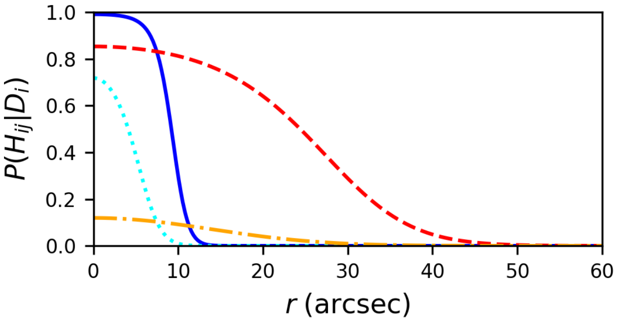


Figure 2.10: Posterior matching probabilities as a function of the angular separation for X-ray sources with different positional uncertainties σ in arcsec and counterparts with varying source densities η in deg^{-2}. Blue solid curve: $\sigma = 3$ and $\eta = 700$; cyan dotted curve: $\sigma = 3$ and $\eta = 30\,000$; red dashed curve: $\sigma = 13$ and $\eta = 700$; orange dashed-dotted curve: $\sigma = 13$ and $\eta = 30\,000$.

are associated to the same object and that the sources are uncorrelated. If the observed properties are more likely for true identifications, the Bayes factor is > 1 and the posterior matching probability increases. On the other hand, the Bayes factor is < 1 and the matching probability is downweighted for properties that are more often observed for random associations. The likelihood of Equation 2.6 is expended to

$$P(D_i|H_{ij}) = \frac{1}{2\pi(\sigma_X^2 + \sigma_O^2)}\, e^{-\frac{r_{ij}^2}{2(\sigma_X^2+\sigma_O^2)}} \cdot B_{ij}. \tag{2.9}$$

Empirically, true and spurious identifications can be approximated by associations with a good positional match and counterparts to shifted sources. However, the details of the estimation of the likelihood functions allow some modifications of personal choice and I discuss in detail the construction of the Bayes map and its influence on the resulting identifications in Sect. 4.3.

Chapter 3

Identification only with geometric matching properties

Some stellar identifications of the XMMSL sources are provided by Saxton et al. (2008). However, most of the stellar counterparts were adopted from the SIMBAD database that combines sources from different catalogs. Therefore, the counterparts are very inhomogeneous and incomplete. In Freund et al. (2018) (Sect. 3.1), my colleagues and I present an identifications of a flux-limited sample of the stellar XMMSL sources.

As shown in Fig. 2.10, the correct identification of an X-ray source can be found only with the geometric matching properties when the positional accuracy of the X-ray sources is high and the counterpart density is low. This is the case for the sources detected in the *XMM-Newton* slew survey (XMMSL; Saxton et al., 2008), which have positional uncertainties of about 8 arcsec. Due to the saturation limit of stellar X-ray sources and the comparatively shallow sensitivity of 5×10^{-13} erg cm^{-2} s^{-1} for stellar XMMSL sources, relatively bright optical counterparts ($G < 15$ mag) are expected for the XMMSL sources (cf. Fig. 1 in Freund et al. (2018) (Sect. 3.1)), and hence, the counterpart density is also low. At the time of submission, only *Gaia* DR1 was available, and due to its incompleteness and missing color information, counterparts from the 2MASS and Tycho2 catalog were added. Statistical matching probabilities are provided for all counterparts, but in contrast to later matching procedures (Freund et al., submitted (Sect. 4.1) and Schneider et al. (2021) (Sect. 5.1)), the estimation of these probabilities is not based on Bayesian statistics because only the geometric match and no further properties are used.

The completeness and reliability of the obtained identifications is significantly improved compared to the classification of Saxton et al. (2008) that is not optimized for stars. Due to the variability of stellar X-ray sources, many of the XMMSL detections are previously unknown X-ray sources.

I programmed the identification procedure, validated the identifications, and analyzed their properties by myself. The coauthors provided valuable ideas to all these steps.

3.1 Publication: The stellar content of the XMM-Newton slew survey

A&A 614, A125 (2018)
https://doi.org/10.1051/0004-6361/201732009

Astronomy & Astrophysics

The stellar content of the *XMM-Newton* slew survey[★]

S. Freund, J. Robrade, P. C. Schneider, and J. H. M. M. Schmitt

Hamburger Sternwarte, Universität Hamburg, 21029 Hamburg, Germany
e-mail: `sebastian.freund@uni-hamburg.de`, `jrobrade@hs.uni-hamburg.de`, `cschneider@hs.uni-hamburg.de`

Received 27 September 2017 / Accepted 14 December 2017

ABSTRACT

Aims. We present a detailed analysis of the stellar content of the current version of the *XMM-Newton* slew survey (XMMSL2).
Methods. Since stars emit only a small fraction of their total luminosity in the X-ray band, the stellar XMMSL2 sources ought to have relatively bright optical counterparts. Therefore the stellar identifications were obtained by an automatic crossmatch of the XMMSL2 catalog with the first *Gaia* data release (*Gaia* DR1), 2MASS, and *Tycho*2 catalogs. The reliability of this procedure was verified by a comparison with the individually classified *Einstein* Observatory medium sensitivity survey X-ray sources and by a crossmatch with the *Chandra* Source Catalog.
Results. We identify 6815 of the 23 252 unique XMMSL2 sources to be stellar sources, while 893 sources are flagged as unreliable. For every counterpart a matching probability is estimated based upon the distance between the XMMSL2 source and the counterpart. Given this matching probability the sample is expected to be reliable to 96.7% and complete to 96.3%. The sample contains stars of all spectral types and luminosity classes, and late-type dwarfs have the largest share. For many stellar sources the fractional contribution of the X-ray band to the total energy output is found above the saturation limit of previous studies ($L_X/L_{\rm bol} = 10^{-3}$), because the XMMSL2 sources are more affected by flares owing to their short exposure times of typically 6 s. A comparison with the second *ROSAT* all-sky survey (2RXS) source catalog shows that about 25% of the stellar XMMSL2 sources are previously unknown X-ray sources. The results of our identification procedure can be accessed via VizieR.

Key words. stars: activity – X-ray: stars – methods: miscellaneous

1. Introduction

Catura et al. (1975) were the first to detect coronal X-ray emission from a star other than the Sun in the bright active binary system Capella, albeit at a level much brighter than typical solar X-ray emission levels. Later, X-ray observations with the *Einstein* Observatory (Vaiana et al. 1981) and then with *ROSAT* showed X-ray emission to be ubiquitous for almost all types of stars (Schmitt et al. 1995; Schmitt 1997; Huensch et al. 1998a,b).

The X-ray properties of stars are usually investigated either by pointed observations of selected X-ray sources – for example, currently with the *XMM-Newton* (Jansen et al. 2001) and *Chandra* space telescopes or via all-sky surveys such as the *ROSAT* all-sky survey (RASS; Voges et al. 1999) or, in the future, the eROSITA all-sky survey (Merloni et al. 2012; Predehl 2017). Such all-sky surveys have the advantage of delivering large samples of X-ray sources that are not biased by the selection of specific sources or specific sky regions. The same applies to the *XMM-Newton* slew survey (XMMSL; Saxton et al. 2008), which delivers data in a spectral range identical to the upcoming eROSITA survey. The XMMSL, however, is somewhat special in the context of X-ray surveys. The *XMM-Newton* satellite also collects X-ray data while slewing from one pointed observation to the next and these data form the basis of the XMMSL, which is regularly updated with the mission.

Naturally, in contrast to true all-sky survey such a survey is very inhomogeneous, but in its current version (XMMSL2) the XMMSL already covers 84% of the sky and includes 29 393 detections of 23 252 unique X-ray sources. We are specifically interested in the stellar content of the XMMSL, hence our task at hand is the development of a procedure that distinguishes stellar sources in the XMMSL from other classes of X-ray emitters such as galaxy clusters and active galactic nuclei (AGN) as reliably and completely as possible. Because of the large number of XMMSL2 sources, this identification process can obviously not be carried out individually by hand, rather an automatic method is required that uses the known properties of stellar (coronal) X-ray sources.

Stellar X-ray sources are relatively faint, when measured in terms of the fractional contribution of the X-ray band to the total energy output, i.e., the $L_X/L_{\rm bol}$-ratio. For example, early-type stars typically satisfy $L_X/L_{\rm bol} \approx 10^{-7}$ (Pallavicini et al. 1981; Berghoefer et al. 1997) and their X-ray emission is generated through radiative instabilities in their radiatively driven stellar winds. In contrast, the X-ray emission observed from late-type, cool stars is produced in hot coronae, and magnetic fields are thought to play a fundamental role for the coronal physics of stars (Pevtsov et al. 2003). The observed X-ray luminosities of late-type stars vary enormously, both in individual cases and in a sample of stars. In the case of flares the X-ray flux can increase by orders of magnitude over timescales of minutes to hours. Also, similar to the solar cycle late-type stars may show modulated X-ray emission on timescales of years related to activity cycles (Hempelmann et al. 2003; Favata et al. 2008; Ayres 2009; Robrade et al. 2012), in addition the X-ray flux of a given star may vary on the timescale of rotation on typically a timescale of a few days and possibly longer. As a class, late-type dwarfs show a rather well-defined maximum fractional X-ray emission

---

[★] Catalog of the stellar XMMSL2 sources is only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via `http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/614/A125` of about $L_X/L_{bol} = 10^{-3}$ during so-called quasi-quiescent periods, i.e., during times without obvious strong flares (Vilhu 1984; Agrawal et al. 1986; Fleming et al. 1988; Pallavicini et al. 1990). A similarly well-defined lower limit does not exist, but Schmitt (1997) showed the existence of a minimum X-ray surface flux of about 10^4 erg s^{-1} cm^{-2} for dwarf stars, which results in a $L_X/L_{bol} \approx 10^{-7}$–$10^{-6}$ for solar analogs.

Stars off the main sequence also show X-ray emission and the X-ray luminosity can be very high, especially for giants that are part of a binary system, for example, RS CVn systems and related systems (Walter et al. 1978; Dempsey et al. 1993). However, little to no X-ray emission is found for red giants beyond the so-called dividing line (Linsky & Haisch 1979; Haisch et al. 1991; Huensch et al. 1996).

The low fractional X-ray luminosity stellar X-ray sources implies that counterparts of these sources ought to be relatively bright in the optical band. Hence, any star, detected for example in the XMMSL, will also be detected in an optical survey of sufficient sensitivity. In this context the currently operating *Gaia* mission (Gaia Collaboration 2016b) is particularly relevant, since *Gaia* will eventually produce a complete optical catalog down to a magnitude of 20 as well as parallaxes, which allows easy distinction of nearby stellar sources from more distant Galactic and extragalactic sources. In November 2016 the first data release of the *Gaia* optical all-sky survey was issued by the Gaia Collaboration (2016a, *Gaia* DR1), and we can therefore start to tap the *Gaia* potential in our effort to identify the stellar XMMSL sources by a crossmatch with the *Gaia* DR1.

The plan of our paper is then as follows. In Sect. 2 we present the properties of the XMMSL and the *Gaia* DR1 catalog along with the complimentary catalogs used in this paper. In Sect. 3 we describe our matching procedure and estimate the expected completeness and reliability of our stellar identification based upon the matching probability of the individual counterparts. We present our results and compare our stellar identifications with those of Saxton et al. (2008) in Sect. 4. Additionally, we test the reliability of our automatic matching procedure by applying it to the Extended Medium-Sensitive Survey (EMSS; Gioia et al. 1990; Stocke et al. 1991) of the *Einstein* Observatory, whose sources have been individually classified by spectroscopy, and by performing a crossmatch with the *Chandra* Source Catalog. In Sect. 5 we compare the X-ray fluxes of the stellar *XMM-Newton* slew survey sources with the corresponding fluxes of the second *ROSAT* all-sky survey (2RXS) source catalog (hereafter RASS catalog, Boller et al. 2016). The properties of the stellar sample of the XMMSL2 sources are presented in Sect. 6 and we draw our conclusions in Sect. 7.

2. Catalog suite

We first provide short descriptions of the various catalogs used in this paper.

2.1. XMMSL catalog

For our stellar identifications, we used the clean version of the XMMSL2 catalog as the X-ray input catalog, which we refer to as XMMSL2 catalog hereafter. For a detailed description of the catalog and its creation, we refer to Saxton et al. (2008), who describe all methods of the production of the first *XMM-Newton* slew survey catalog in detail; these methods are very similar to those of the XMMSL2 catalog. Briefly, this catalog contains detections with a detection likelihood of $DET\_ML > 10.5$ in general and of $DET\_ML > 15.5$ for sources with higher than the usual background. The positional accuracy is typically about 8 arcsec and, according to Saxton et al. (2008), about 4% of the sources detected in the full band (as well as 0.7% and 9% of the sources detected in the soft and hard band, respectively) are spurious.

All *XMM-Newton* slews are treated individually during the creation of the XMMSL2 catalog, such that every X-ray detection leads to a new entry in the catalog, even if the same source has been detected in a previous slew. In a second step, detections lying within 30 arcsec in different slews are then considered to be multiple detections of the same source and are given the same source name. Therefore, the 29 393 XMMSL2 detections actually come from 23 252 unique X-ray sources. For our identification of stars, we only used the unique XMMSL2 sources and coordinates of the detection with the highest detection likelihood. Further, we used the median X-ray flux for sources with multiple detections.

In the XMMSL2 catalog the source count rates are given in three different energy bands, i.e., the total band (0.2–12 keV), soft band (0.2–2 keV), and hard band (2–12 keV).

2.2. Gaia DR1

The *Gaia* DR1 catalog contains the positions and *G* band magnitudes of 1.1 billion sources. For a subset of 2 million stars, parallaxes and proper motions were calculated from information provided by the HIPPARCOS and *Tycho*2 catalogs; this subset is called the *Tycho-Gaia* astrometric solution (TGAS). The positional and photometric uncertainties of all catalog entries are negligible compared to the uncertainties in the X-ray data (better than 10 mas and 0.03 mag, respectively).

Unfortunately, the *Gaia* DR1 catalog has a only preliminary character (Gaia Collaboration 2016a), and for our purpose, the relevant known limitations are its incompleteness for very bright sources $\lesssim$7 mag, sources with high proper motion, extremely blue or red sources, and sources located in dense areas on the sky and or in binary systems. To overcome these limitations and to obtain color information for spectral type classification, we considered complementary catalogs.

2.3. Complementary catalogs

To obtain colors and hence spectral types for the X-ray counterparts, we used the 2MASS catalog (Skrutskie et al. 2006), i.e., an infrared catalog that is particularly useful for stars of late spectral type (cf. Fig. 1), and the *Tycho*2 catalog (Høg et al. 2000) to identify the brighter sources. In addition, we considered information provided by the Bright Star catalog (Hoffleit & Jaschek 1991) and the catalog by Lépine & Gaidos (2011) of bright M dwarfs (Lepine catalog). In the following we describe the matching procedure only for the *Gaia* DR1, 2MASS, and *Tycho*2 catalogs, which provide the vast majority of our stellar identifications, but the presented method was also applied to the Bright-Star and the Lepine catalog and appropriate matching distances and probabilities were estimated for these catalogs.

We expect essentially all stellar XMMSL2 sources to have a 2MASS counterpart. Since the completeness of the 2MASS catalog is >99%, our procedure is not influenced by the incompleteness of the catalogs used for the identification.

3. Data analysis

Our matching procedure is based upon the angular distance between the XMMSL2 X-ray sources and potential stellar catalog counterparts, where we corrected the position of the stellar counterpart for proper motion, if the proper motion is given in the catalog. This procedure includes optical and near-infrared (NIR) brightness cuts to limit the number of chance alignments with faint sources that are unlikely to be responsible for X-ray emission. The appropriate magnitude cuts depend on the X-ray detection limit, i.e., the conversion between observed count rate and flux, and the expected ratio between X-ray and optical or NIR fluxes. We then tested our procedure against random sources, motivate the used matching distances, and discuss the further applied selection procedures; a flow chart of our matching procedure is given in Appendix A.

3.1. X-ray fluxes

The conversion between the measured count rates and the derived X-ray fluxes depends on the spectral model assumed for the X-ray source. Therefore, we did not adopt the X-ray fluxes given in the XMMSL2 catalog, which are estimated by applying a spectral model typical for AGN; instead, we used our own conversion by adopting a spectral model that is appropriate for stellar X-ray sources. Specifically, we assumed optically thin emission and adopted an APEC thermal plasma model with a temperature of 5×10^6 K and solar metallicity. We neglected interstellar extinction because we expected to find most of the stellar counterparts within 150 pc. However, a few sources located in star forming regions might be affected by the interstellar absorption, but with the data at hand we cannot identify these sources. Furthermore, we converted these fluxes into the *ROSAT* band to compare the XMMSL2 X-ray fluxes with previous measurements and used a count rate [cts s^{-1}] to flux conversion factor of 1.24×10^{-12} erg cm^{-2} s^{-1} for the soft and total band. The conversion factor is relatively insensitive to the assumed temperature; for the 2–35$\times10^6$ K range, it changes by 6% and 15% for the total and soft band, respectively. The hard band is generally less suitable to observe stellar X-ray sources because coronal X-ray sources are typically rather soft X-ray emitters, unless they are heavily absorbed and the effective area of the *XMM-Newton* decreases for high energy photons. We formally adopted a flux conversion factor of 5.93×10^{-10} erg cm^{-2} s^{-1} per count rate of 1 cts s^{-1} for the hard band, but this value is – naturally – very sensitive to the assumed model temperature, and we do not expect stellar sources to be detected only in the hard band.

The detection limit of the XMMSL2 catalog is typically ~0.4 cts s^{-1} for a source passing through the center of the detector at a typical background level. With the adopted conversion factor this corresponds to an X-ray flux of 5×10^{-13} erg cm^{-2} s^{-1}, which we used to derive the minimum optical brightness of potential stellar counterparts.

3.2. Minimal optical brightness of the stellar XMMSL2 sources and magnitude cutoff

With our estimate of the limiting XMMSL2 flux and the saturation limit of stellar X-ray emission, we can compute the minimal bolometric flux of a possible stellar counterpart to an XMMSL2 X-ray source. Given this minimal bolometric flux, we used Table 3 of Worthey & Lee (2011) to calculate the minimum optical brightness in different photometric bands as a function of the effective temperature (cf. Sect. 3.6), again neglecting interstellar absorption, and show the computed magnitudes in the V, G, and J bands versus the effective temperature in Fig. 1. As is clear from Fig. 1, all stellar XMMSL2 sources ought to be relatively bright in the optical with $G \lesssim 14$ mag. Furthermore, the stellar sources are also bright in the infrared band with $J \lesssim 12$ mag; in particular, very late spectral type dwarfs can be faint in the V and G bands, but should still be bright in the J band of 2MASS.

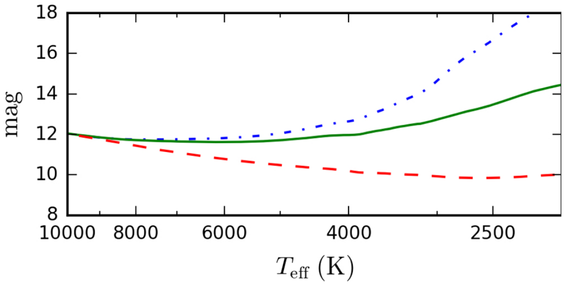


Fig. 1. Apparent magnitude in different photometric bands for a star with saturated X-ray emission at the detection limit of the XMMSL2 catalog vs effective temperature. The blue dash-dotted line indicates the V band; the green solid line indicates the G band; and the red dashed line indicates the J band.

In practice, the magnitude limit shown in Fig. 1 is not sharp. First of all, there appears to be some intrinsic scatter in the saturation limit (Pizzolato et al. 2003) and, second, sources might be caught during an X-ray flare during the slew survey observations and therefore produce more X-ray flux than expected. Hence, the $L_X/L_{\rm bol}$ ratio of coronal sources detected in the XMMSL2 could be higher than the saturation limit and their optical brightness lower than the predictions shown in Fig. 1. For example, Stelzer et al. (2006) reported a flare with a peak X-ray luminosity 200–300 times above the quiescence emission and with an increase in optical brightness of $\Delta V = 6$ mag for the star LP 412-31; however, these extreme flare events are rare and it is unlikely that *XMM-Newton* slews over a star during the peak of such an extreme flare.

Yet to allow for some margin in these cases, we adopted a magnitude cutoff at $G = 16$ mag for *Gaia* sources, at $J = 12$ mag for 2MASS sources and no cutoff for *Tycho*2 sources; however, if neither a *Tycho*2 nor a 2MASS counterpart is found, the *Gaia* cutoff is set to $G = 15$ mag. These cutoff values are clearly sufficient to find all potential stellar counterparts emitting at the X-ray saturation level. The achieved stellar activity margin depends on spectral type, for example, for stars with an effective temperature of 3000 K (spT: M5V) the X-ray flux limit is about $\log(F_X/F_{\rm bol}) < -2.2$ at minimum optical brightness. We remark in passing that the adopted magnitude cutoffs are well above the completeness limits of the *Gaia* and 2MASS catalog, such that all stellar XMMSL2 sources ought to be included in the catalogs.

3.3. Random matches and magnitude distribution

With its 1.1 billion sources the mean distance between two *Gaia* entries is about 20 arcsec, thus finding a *Gaia* catalog entry in the vicinity of an XMMSL2 source is not surprising. In order to investigate the influence of random coincidences on our matching procedures, we carried out *Gaia* identifications with randomly generated X-ray sources. Since both the XMMSL2 catalog and the stellar catalogs chosen for matching are nonuniform, it is important to preserve the global spatial distribution of the X-ray sources in the randomly generated X-ray samples; this is achieved by using all cataloged XMMSL2 sources, but shifting their positions uniformly between a distance of 240 arcsec and 1200 arcsec along a randomly chosen direction.

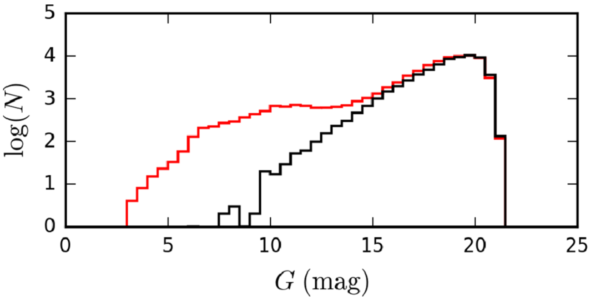


Fig. 2. Magnitude distribution of the *Gaia* counterparts. The red line indicates the magnitude distribution of the *Gaia* associations of the real XMMSL2 sources; and the black line indicates the magnitude distribution of the *Gaia* associations with the randomly generated sources

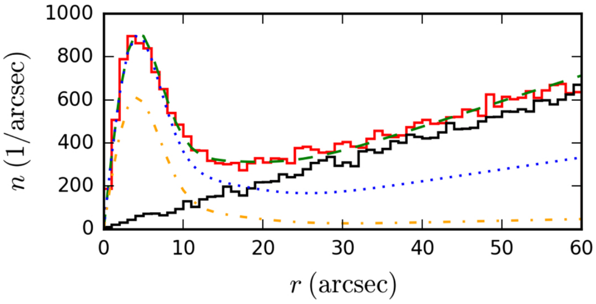


Fig. 3. Distribution of the distances between the XMMSL2 sources and the stellar counterparts. The red solid line represents a histogram of the distances between the XMMSL2 sources and the *Gaia* counterparts. The black solid line represents a histogram of the distances between the randomly generated sources and the *Gaia* counterparts. The dashed and dotted curves represent the best fits of Eq. (1) for the different catalogs with the parameters of Table 1. The green dashed curve represents the *Gaia* catalog, the blue dotted curve represents the 2MASS catalog, and the orange dash-dotted curve represents the *Tycho*2 catalog.

In Fig. 2 we show the thus obtained G magnitude distribution of all *Gaia* matches (using a matching distance of 20 arcsec to the XMMSL2 sources) as well as that of the randomly generated XMMSL2 sources. Obviously, these distributions substantially differ from each other. While the *Gaia* magnitude distribution of the true XMMSL2 sources is bimodal with a broad first peak near $G \approx 10$ mag, the *Gaia* magnitude distribution of the random XMMSL2 sources steadily increases up to the magnitude cutoff of the *Gaia* catalog near $G \approx 20$ mag. Interestingly, the identified number of true XMMSL2 sources exceeds that of randomly generated XMMSL2 sources up to $G \approx 20$ mag. Thus, even at faint magnitudes some of the *Gaia* counterparts appear to be the correct, albeit not necessarily stellar, identifications. However, it is also clear that for a magnitude $G \approx 14$ mag, the chance to obtain a random match exceeds 50% using the matching distance of 20 arcsec.

3.4. Choice of the matching distance

Next, we considered the (differential) number of matched XMMSL2-*Gaia* sources (choosing only *Gaia* entries with $G < 15$ mag) as a function of matching distance and show the resulting histograms for the real XMMSL2 sources and the randomly generated XMMSL2 sources in Fig. 3. Again, the two distributions differ substantially. The distribution of the randomly generated sources increases linearly as expected, while the distribution of the real XMMSL2 sources is bimodal. At small matching distances it is dominated by a Gaussian-type distribution up to a distance of 15 arcsec. We find this distribution to be better fitted by a double Gaussian than by a single Gaussian distribution for the XMMSL2 sources. However, this is only an empirical description without any deeper physical meaning. At larger distances the distribution of the real XMMSL2 sources approximates the linear distribution of the randomly generated sources. The peak at small distances contains the true matches, while the linearly increasing population of matches represents random associations. In Fig. 3 we also plot the corresponding distributions resulting from the 2MASS catalog using a magnitude limit of $J = 12$ mag and the full *Tycho*2 catalog. These distributions are qualitatively similar to the *Gaia* distribution and differ only quantitatively because of the smaller number of catalog entries.

3.4.1. Matching probability

The distribution of the real XMMSL2 sources can be well fitted with a double Gaussian (describing the uncertainty in the XMMSL2 positions) and a linear curve (describing the random matches) using the ansatz

Table 1. Fitted parameters of Eq. (1).

	Gaia	2MASS	*Tycho*2
A [1/arcsec^2]	310 ± 6	318 ± 6	233 ± 5
B [1/arcsec^2]	29 ± 1	27 ± 1	11.7 ± 0.6
M [1/arcsec^2]	11.89 ± 0.09	5.54 ± 0.06	0.77 ± 0.03

$$n(r) = A \cdot r \cdot \exp\left(-\frac{r^2}{2\sigma_1^2}\right) + B \cdot r \cdot \exp\left(-\frac{r^2}{2\sigma_2^2}\right) + M \cdot r. \qquad (1)$$

In Eq. (1) the parameters σ_1 and σ_2 are the standard deviations of the Gaussian distributions, which are independent of the matching catalog; we find $\sigma_1 = 4.0$ arcsec and $\sigma_2 = 9.9$ arcsec. The values for the parameters A, B, and M depend on the source densities of the catalogs. In Table 1 we provide the best fit parameters for the catalogs used in this paper, and in Fig. 3 we give a visual representation of the best fit curves.

With the fitted parameters of Eq. (1) the probability p of a match at the distance r to be the true counterpart can be estimated through the expression

$$p(r) = \frac{A \cdot r \cdot \exp\left(-\frac{r^2}{2\sigma_1^2}\right) + B \cdot r \cdot \exp\left(-\frac{r^2}{2\sigma_2^2}\right)}{A \cdot r \cdot \exp\left(-\frac{r^2}{2\sigma_1^2}\right) + B \cdot r \cdot \exp\left(-\frac{r^2}{2\sigma_2^2}\right) + M \cdot r}. \qquad (2)$$

Although the standard deviations σ_1 and σ_2 take the same value for all matching catalogs, the probability varies for a specific matching distance because for a catalog with a lower source density one is less likely to find a random association. The differences in the source densities are caused by the magnitude ranges covered by the various catalogs. However, the matching probability does not consider the different brightnesses within a catalog. Therefore, the fraction of spurious identifications ought to be larger near the magnitude cutoff.

3.4.2. Missed and spurious identifications

The number of stellar identifications of the XMMSL2 sources missed in our sample, i.e., the completeness, and the number of spurious identifications in the sample, i.e., the reliability, depend on the chosen matching criteria. Especially, the probability cutoff above which we assume matches to be the true counterpart controls these two characteristics (cf. Sect. 3.4.1). The number of spurious identifications N_{spurious} in the sample with a constant matching distance r can be estimated through

$$N_{\text{spurious}} = \int_0^r M r' \mathrm{d}r'. \tag{3}$$

In our sample the matching distance is not constant, but depends on the catalog used for the identification. The differential matching probability, defined in Eq. (2), gives the probability for a match at the distance r to be the true counterpart. Hence we calculated the probabilities p_i of each match up to a distance of 40 arcsec where the probability to be the true counterpart drops to below 1% for all our catalogs, i.e., these matches are negligible.

Given the probabilities p_i and the chosen probability cutoff, the number of spuriously identified XMMSL2 sources can be estimated by summing the inverse probabilities of all sources with a probability higher than the cutoff $N_{>\text{cutoff}}$. For a XMMSL2 source with N_{matches} matches, the probability that none of the matches is the true counterpart is given by the product of the inverse probabilities of all matches. Hence, we estimated the number of spuriously identified sources through

$$N_{\text{spurious}} = \sum_i^{N_{>\text{cutoff}}} \prod_j^{N_{\text{matches},i}} \left(1 - p_j\right) \tag{4}$$

and we defined the reliability r of the matched sample as

$$r = \frac{N_{>\text{cutoff}} - N_{\text{spurious}}}{N_{>\text{cutoff}}}. \tag{5}$$

We estimated the number of missed identifications by summing the probabilities of all matches with a probability lower than the cutoff $N_{<\text{cutoff}}$

$$N_{\text{missed}} = \sum_i^{N_{<\text{cutoff}}} p_i. \tag{6}$$

and define the completeness C as

$$C = \frac{N_{>\text{cutoff}}}{N_{>\text{cutoff}} + N_{\text{missed}}}. \tag{7}$$

3.5. Associations in multiple catalogs

A few XMMSL2 sources have counterparts in several catalogs, therefore we must determine whether we are considering the same counterpart or not. In some catalogs the identifications of other catalogs are specified; for example, the TGAS sample of the *Gaia* DR1 contains the *Tycho*2 identifier, which is, in this specific case, based exclusively on apparent sky distance. If no identifier is specified, we associated counterparts if their distance is smaller than 1 arcsec or their extrapolated V band magnitude difference is smaller than 1.5 mag and their distance is smaller than 4 arcsec. The exact distances do not influence the result significantly. We chose the closest match if these conditions are met by multiple counterparts. For measurements given in multiple catalogs, we used the value of the catalog with the highest accuracy of the respective measurement. To estimate the probability of a random match, we used the catalog with the highest probability.

3.6. Additional stellar properties

With the magnitudes in the available photometric bands, we estimated the effective temperatures, the bolometric magnitudes and fluxes and the V band magnitudes for the sources with a 2MASS and *Gaia* counterpart only. We adopted the relations in Table 3 of Worthey & Lee (2011), applying solar metallicity and a surface gravity of $g = 10^{4.5}$ cm s^{-2} corresponding to dwarfs and used a linear interpolation of the values given in the table. We did not perform any corrections for extinction. Since Worthey & Lee (2011) provided the colors in the photometric system of Bessel & Brett, we applied the relation given by Carpenter (2001) and in Explanatory Supplement to the 2MASS All Sky Data Release and Extended Mission Products[1] to obtain the 2MASS colors. We further adopted the correlation to the *Gaia* band from Jordi et al. (2010) and the *Gaia* Data Release Documentation[2].

For the calculation we adopted the $V-K$ color for the sources with a counterpart in the *Tycho*2 catalog and the 2MASS catalog, while we used the $B-V$ color and the $J-K$ color for the sources with a counterpart only in the *Tycho*2 catalog or the 2MASS catalog, respectively. We could not estimate the bolometric flux and the effective temperature for a few sources (~2%) because they have a *Gaia* counterpart only or they are extremely red and lie outside the region defined in Table 3 of Worthey & Lee (2011).

We find a trigonometric parallax for 57% of the stellar XMMSL2 sources and for these sources, we further estimate the distance, the V band and bolometric absolute magnitude, and the bolometric and the X-ray luminosity.

3.7. Cleaning procedures

In our catalog we introduce different XMMSL2 source flags, if the sources have measurements of low quality or to indicate likely nonstellar objects. A few object classes are identified as stellar sources by the procedures outlined above, which are not the focus of this work; examples are high- and low mass X-ray binaries, where the X-ray emission is not predominantly produced by the star, but by matter accreted onto a compact object. Many of these objects were already excluded because they generally have faint optical counterparts due to their very large X-ray/optical flux ratios, but we additionally excluded sources that have a known accreting object in the SIMBAD database (Wenger et al. 2000) within a distance of 30 arcsec to the XMMSL2 detection. Furthermore, we flagged sources that have a known galaxy cluster within 60 arcsec or an AGN located within 30 arcsec as listed the SIMBAD database. For such X-ray sources, the stellar object and extragalactic object are both plausible counterparts, given the available information.

We additionally flagged all sources that are not detected in all 2MASS bands and that have no association in other catalogs. Furthermore, the stellar identifications are unreliable if they are flagged as extended in the 2MASS catalog or if they are

---

[1] `http://www.ipac.caltech.edu/2mass/releases/allsky/doc/explsup.html`

[2] `https://gaia.esac.esa.int/documentation/GDR1/`

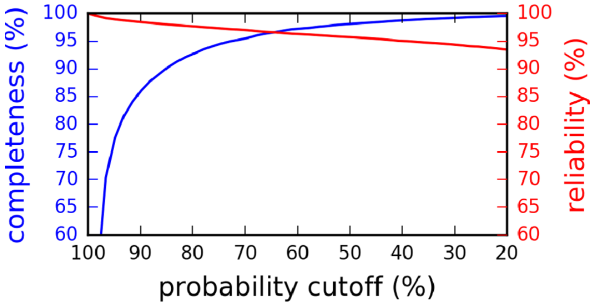


Fig. 4. Completeness (blue curve) and reliability (red curve) of the sample as a function of the probability cutoff.

detected in the XMMSL2 hard band only. We expected sources to be affected by optical loading (i.e., X-ray events created by an excess of optical photons in the pn camera) if they are flagged in the XMMSL2 catalog and do not have a RASS counterpart (cf. Sect. 5). Some sources have erroneous 2MASS photometry or an unusual color and very likely the derived magnitudes and stellar properties are unreliable. Additionally, we flagged sources with a high X-ray to the bolometric flux ratio. Specifically, we flagged sources with a high $F_X/F_{\rm bol}$ ratio only in the slew survey and set an additional flag for sources that consistently have a high $F_X/F_{\rm bol}$ ratio in both the slew and the RASS surveys and another flag is used for sources without RASS counterpart.

All used flags are summarized in Appendices A and B. We generally excluded flagged sources from our subsequent analysis, but discuss some of their properties in the following sections.

4. Results

4.1. Stellar counterparts, completeness, and reliability

Figure 4 shows the completeness and reliability of the sample and its dependence on the matching probability for the non-flagged sources. To balance completeness and reliability, we choose a matching probability >2/3, i.e., about intersection of the two curves, to derive the stellar catalog of XMMSL2 sources and obtain a completeness of 96.3% (Eq. (6)) and a reliability of 96.7% (Eq. (4)) ignoring sources that are flagged by our cleaning procedure in the calculations. The matching radii correspond to 10.8 arcsec, 13.9 arcsec, and 19.9 arcsec for sources with a counterpart in the *Gaia*, 2MASS and *Tycho*2 catalogs, respectively.

Combing these matching radii with our adopted magnitude cutoffs, we find at least one stellar counterpart for 6815 of the 23 252 XMMSL2 sources. Our cleaning procedures reduce the sample to 5920 sources, implying that 25.5% of the XMMSL2 sources are stellar sources.

4.2. Single and multiple counterparts

We specify the number of selected stellar counterparts per XMMSL2 source in Table 2. For the 5042 XMMSL2 sources with a single counterpart we present the histogram of the distances between the XMMSL2 sources and the counterparts in Fig. 5. No strong discontinuities are visible at the adopted matching distances, only at ≈14 arcsec, i.e., for 2MASS identification without a *Tycho*2 counterpart, a slight drop is visible. Sources with a distance > arcsec have a *Tycho*2 identification and the few sources at a distance larger than 20 arcsec have a counterpart in the BrightStar or Lepine catalog that we include because of the small source density of these catalogs.

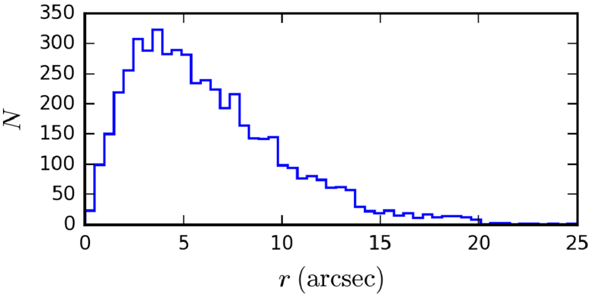


Fig. 5. Distances between the XMMSL2 sources and the counterparts for the 5042 XMMSL2 sources with a single match.

Table 2. Number of multiple counterparts.

Number of stellar counterparts	Number of sources
1	5042
2	761
3	103
>3	14

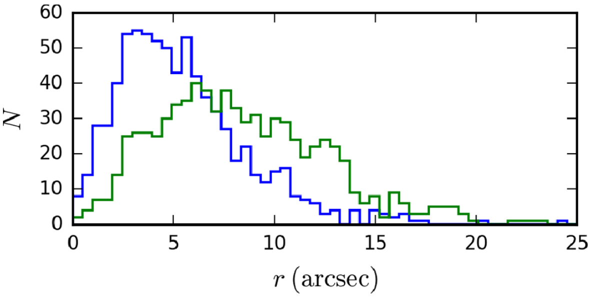


Fig. 6. Distances between the XMMSL2 sources and the counterparts for the 761 XMMSL2 sources with two matches. The blue curve shows the distance of the sources with the higher matching probability; the green curve of the sources with the lower matching probability.

We find that roughly 15% of the stellar XMMSL2 sources have more than one plausible stellar counterpart. In Fig. 6 we show the distances between the XMMSL2 sources and the counterparts for the 761 XMMSL2 sources with two matches. The mean angular separation of the counterparts with the higher matching probability for each XMMSL2 sources is smaller, but overall the two distributions are quite similar, i.e., both distributions have a maximum at small distances and the number of sources decreases for larger distances. In Fig. 7 we show the histogram of the angle between the XMMSL2 source and the counterparts for the sources with two counterparts. The distribution has a maximum at 180°, which implies that the XMMSL2 source tends to lie between both candidates. Thus, the cataloged XMMSL2 source is often likely to be a combination of the X-ray emissions of the two sources. We ignore sources with multiple counterparts when investigating the properties of the stellar counterparts.

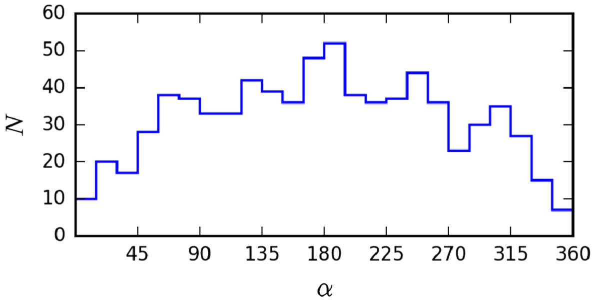


Fig. 7. Distribution of the angle between the XMMSL2 source and the counterpart for the 761 sources with two counterparts.

Table 3. Catalogs of the counterparts.

Catalog	N	Catalog	N
TMG	3595	M	116
MG	1784	T	54
TM	318	G	15
TG	36	other	2

Notes. G: *Gaia*, M: 2MASS, T: *Tycho*2.

4.3. Catalogs of the counterparts

In Table 3 we denote the catalogs providing of the most probable counterpart. Most counterparts have a *Gaia* and 2MASS identification as expected with 490 counterparts missing in the *Gaia* DR1 catalog. This is likely caused by the known incompleteness of the *Gaia* DR1 catalog. Furthermore, some of the *Gaia* sources cannot be identified with their 2MASS counterparts, possibly because of observing epochs that differ by about 15 years and unknown proper motion; 107 counterparts do not have a 2MASS counterpart. Multiple stars that are resolved in the *Tycho*2 or *Gaia* catalog, but not in 2MASS, could explain most of the 107 sources that do not have a 2MASS counterpart. The two sources denoted as “other” in Table 3 only have a counterpart in the Lepine or BrightStar catalog with extremely high proper motions (>4 arcsec yr^{-1}) and therefore, the 2MASS and *Gaia* counterparts, having no proper motion, lie outside of our initial matching radius of 40 arcsec (cf. Appendix A).

4.4. Comparison to the identifications of Saxton et al. (2008)

The XMMSL2 catalog provides identifications and classifications for about 70% of the X-ray sources as a result of a crossmatch with the SIMBAD, NED, and other databases and catalogs. However for some sources the classification only contains the region of the electromagnetic spectrum in which the source has been detected, for example, “X-ray”, and provides little or no insight into the physical nature of the source. The classification adopted in the XMMSL2 catalog Saxton et al. (2008) uses different catalog resources and enables an independent comparison with our results.

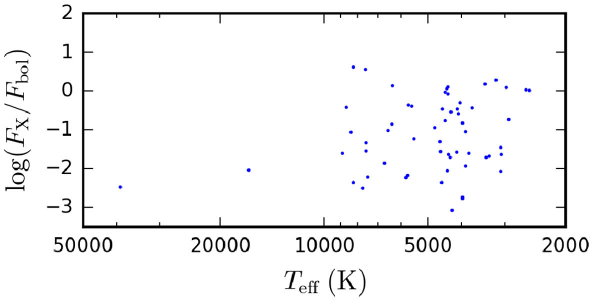


Fig. 8. $F_X/F_{\rm bol}$ distribution as a function of effective temperature of the sources classified as stellar in the XMMSL2 catalog, but missing a stellar counterpart with our approach.

There is large overlap in the identification with 4231 sources consistently classified as stellar; the XMMSL2 catalog identifies a total of 5094 sources as stars. Our stellar identification was not confirmed for 1689 sources, however the vast majority (1671) of these have either no identification or the classification contains only the region of the electromagnetic spectrum. The XMMSL2 contradicts our stellar identification only for 18 (0.4%) sources, typically referring to an additional plausible counterpart.

On the other hand, 863 sources are classified as stars in the XMMSL2 catalog, in which we did not find a stellar counterpart satisfying our selection criteria. Out of these, 531 were excluded by our cleaning procedures, which argues against a true stellar identification. To validate the reliability of the remaining 332 stellar identifications, we inspected the $\log F_X/F_{\rm bol}$ ratio as a function of the effective temperature for the 59 sources for which colors from the SIMBAD database are available (see Fig. 8). For most of these sources, we find $\log F_X/F_{\rm bol} > -2$ or $T_{\rm eff} > 10\,000$ K and $\log F_X/F_{\rm bol} > -3$, i.e., highly unlikely values for stellar X-ray sources. Overall, only a few of these sources remain as plausible stellar counterparts that are missed due to our chosen magnitude cutoff. Based on this fraction, we estimate that about 1% of the stellar counterparts are missed by our procedure due to the applied brightness limit.

4.5. Validation of our procedure

4.5.1. Application of our procedure to the EMSS catalog

We validated the reliability of our identification approach by applying our procedures to the Extended Medium-Sensitive Survey (EMSS; Gioia et al. 1990; Stocke et al. 1991). The EMSS catalog contains 835 X-ray sources detected in 1435 pointings with the Imaging Proportional Counter (IPC) on board the *Einstein* Observatory at high galactic latitudes ($|b| > 20°$). More than 96% of the EMSS sources have been individually identified with spectroscopically classified counterparts. The reliability of these identifications should be very high, therefore the EMSS catalog is an ideal tool to test the reliability of our automatic procedure.

To identify the stellar content of the EMSS, we used basically the same procedure as for the XMMSL2 catalog, but we find a single Gaussian plus linear curve to describe adequately the matching distance distribution of the EMSS sources. Furthermore, we searched in the SIMBAD database for accreting and extragalactic objects within 50 arcsec, which is the typical 90% confidence error circle radius of the EMSS sources.

We identified 210 of the 835 EMSS sources as stellar, whereas we expected about 15 identifications to be spurious and a similar number to be missed by the probability cutoff. These numbers can be compared to the identifications of the EMSS catalog, which contain 217 stars; 192 sources are consistently classified by the EMSS catalog and by our procedure.

The EMSS catalog thus identifies 25 sources as stars that we did not find with our procedure as stellar sources. One of

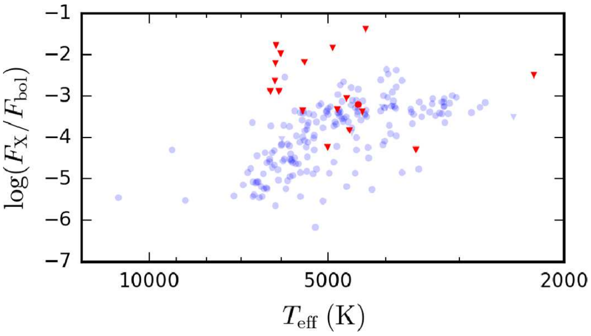


Fig. 9. F_X/F_{bol} distribution of the EMSS sources as a function of effective temperature. The blue symbols represent sources consistently identified as stars, and red symbols show sources classified as stars by our procedure, but with contradicting EMSS classification. The triangles denote possible extragalactic sources.

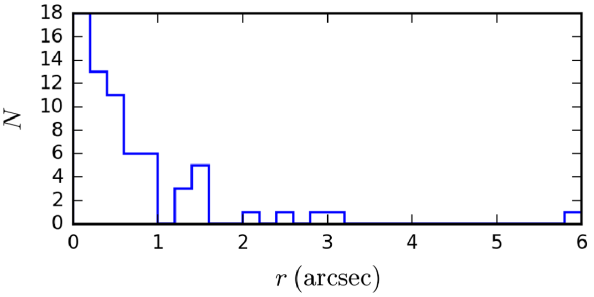


Fig. 10. Angular separation between the stellar identification and the closest CSC counterpart for the XMMSL2 sources with a single match.

these sources is marked as a white dwarf and 7 are marked as cataclysmic variables. Hence, 17 sources are stellar X-ray sources that are missed by our procedure, 10 because of the large distance to the EMSS source, 6 because they have magnitudes of $J > 12$ mag, and one case is uncertain. We applied the same magnitude cutoff to the EMSS catalog as for the XMMSL2 catalog, which is a simplification and not an optimal cutoff for all EMSS sources. Hence, slightly fewer sources are missed because of the probability cutoff than expected.

For 18 sources the EMSS classification explicitly contradicts our stellar identification. However, for most of these sources there are two plausible counterparts in the SIMBAD database, i.e., one stellar counterpart and one extragalactic counterpart. So the stellar identifications may be random associations, which is supported by the number of flagged sources in the total sample (23/210) compared to consistently identified stars (6/192). The number of 18 random associations is comparable to the expectation of 15 spurious identifications.

In Fig. 9 we present the $\log F_X/F_{bol}$ distribution of the 210 EMSS sources that we identify as stellar sources. Obviously, about half of the sources that have a contradicting classification in the EMSS catalog lie above the distribution of the sources that are consistently classified as stars. The other are at least plausible stellar X-ray sources, but we find that they have generally a high angular separation and further they are mostly flagged as having a plausible extragalactic counterpart. Therefore, there is a high chance that these stellar counterparts are random associations.

Yet, in summary the completeness and reliability of our automatic identifications reach the expectations. Therefore we are confident that the reliability and completeness calculated in Sect. 3.4.2 represent the true reliability and completeness of the stellar XMMSL2 sample.

4.5.2. Matching with the *Chandra Source Catalog*

In comparison to *XMM-Newton*, the *Chandra* X-ray Observatory (CXO; Weisskopf et al. 2000, 2002) provides more accurate positions of the X-ray sources owing to its sub-arcsec on-axis point spread function (PSF). However, in its current release 1.1 the *Chandra* Source Catalog (CSC; Evans et al. 2010) covers less than 1% of the sky. Nevertheless, the precise positions of the CSC sources give us the opportunity to validate some of our stellar identifications. Therefore, we performed a crossmatch of our stellar identifications with the CSC, applying a matching distance of 60 arcsec. Although the CSC sources generally have highly accurate positions, for some sources the positional uncertainty is much larger, and hence, we excluded sources with a 95% confidence error circle radius larger than 5 arcsec from our analysis.

In this fashion we find a CSC counterpart for 94 of the 5920 stellar identifications and show the angular separation between the stellar source and the CSC counterpart for the XMMSL2 sources with a single stellar identification in Fig. 10. For most of the sources (86) the distance between the CSC source and our best stellar identification is <2 arcsec as expected for the CSC sources or a larger angular separation can be explained by an unusual high positional uncertainty of the CSC source. For four sources the position of the CSC source indicates that our second best stellar counterpart is the correct identification of the X-ray source, and in four cases none of our stellar identifications lie within the 95% confidence error circle of the CSC source.

In summary, the crossmatch with the CSC confirms at least one of our stellar sources for 95.7% of the XMMSL2 sources, which is in good agreement with the reliability calculated in Sect. 3.4.2.

4.6. Catalog release

We released the catalog of the stellar XMMSL2 sources at VizieR. This catalog contains our stellar identifications with a matching probability >2/3. While we discuss in this paper only the properties of the unflagged sources, the released catalog also includes the stellar counterparts that are flagged by our cleaning procedure. The XMMSL2 sources with multiple stellar counterparts have multiple entries in our catalog, one entry for each counterpart. We describe all new columns of our catalog in Appendix B.

5. RASS counterparts

Before turning to the physical properties of our stellar counterparts, we compared and crossmatched the results from the XMMSL2 to the second RASS-catalog. Since the RASS flux limit is deeper ($\sim 2\times10^{-13}$ erg cm^{-2} s^{-1}) than that of the *XMM-Newton* slew survey ($\sim 5\times10^{-13}$ erg cm^{-2} s^{-1}), one would naively expect that all XMMSL2 sources should have RASS counterparts. However, we found a RASS counterpart within 60 arcsec for only 75.2% of the sources; the matching fraction increases for multiple detected XMMSL2 sources to 91.6%.

To investigate the properties of the stellar XMMSL2 sources without a RASS identification, we compared their distribution as a function of the apparent bolometric magnitude and X-ray activity level to the full sample. As shown in Fig. 11, the fraction of

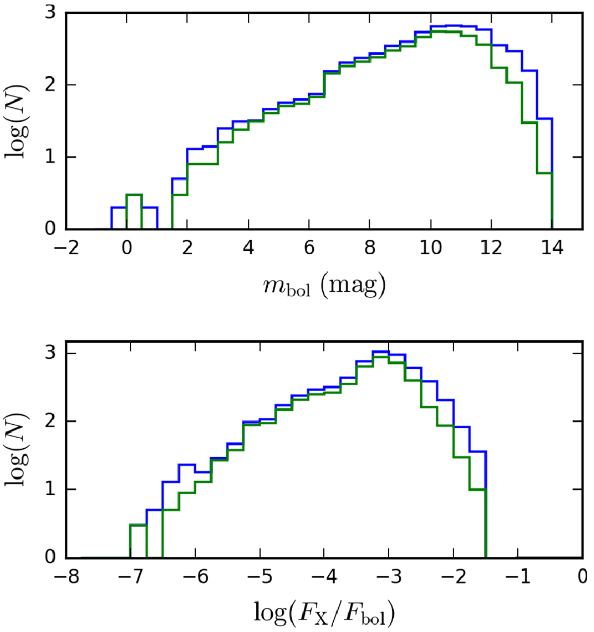


Fig. 11. Number of all stellar XMMSL2 sources (blue) and of those with a RASS counterpart (green) as a function of (*top*) apparent bolometric magnitude and (*bottom*) X-ray activity.

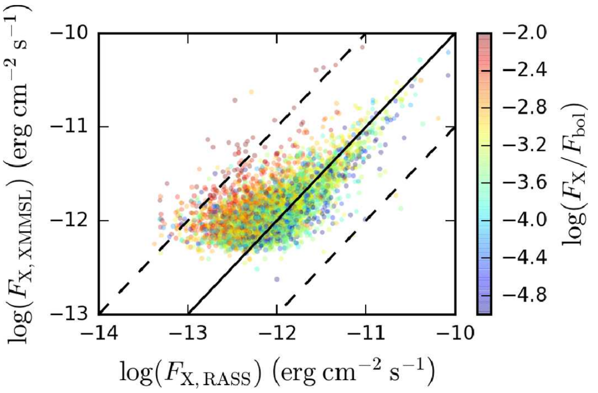


Fig. 12. Comparison of the X-ray fluxes measured by *XMM-Newton* and *ROSAT*. Solid line indicates the same X-ray flux in the XMMSL2 and RASS catalogs, dashed lines indicate flux difference of a factor of ten. The color scales with the stellar X-ray activity measured by *XMM-Newton*.

sources with a RASS identification are reduced for two different types of sources: first for bright sources and second for highly active sources. For high activity stars we expect frequent flaring; thus, the small fraction of RASS identification is caused by the intrinsic variability of these stars.

At bright magnitudes we expected the small fraction of RASS identifications to be due to optical loading; 102 of the stellar XMMSL2 sources are flagged in the XMMSL2 catalog as possibly affected by optical loading. However, 68 of these sources have a RASS identification and, hence, should be considered to be true X-ray emitters; RASS data are not affected by optical contamination. Furthermore, 6 sources brighter than 5 mag without a RASS counterpart are not flagged. Therefore we conclude that the influence of optical loading cannot be reliably determined with the slew data only and hence, the optical loading flag should be used as an indicator together with additional information about the source. The absolute number of XMMSL2 sources affected by optical loading is small and, therefore, intrinsic variability is the main reason why many stellar XMMSL2 sources do not have a RASS identification. We flagged those stellar XMMSL2 sources that are marked as possibly affected by optical loading and that do not have a RASS identification (cf. Sect. 3.7).

For the XMMSL2 sources with a RASS counterpart two independent X-ray flux measurements are available. We converted the measured count rates of the XMMSL2 and RASS sources by applying the conversion factors defined in Sect. 3.1 and by Schmitt et al. (1995), respectively. In Fig. 12 we compare the X-ray fluxes measured by *XMM-Newton* and *ROSAT* for the stellar XMMSL2 sources. For most sources the XMMSL2 flux is higher than the RASS flux and the median flux ratio is 1.4. The difference generally increases with increasing X-ray activity of the star as indicated by the color coding of Fig. 12. For many XMMSL2 sources only 4–5 X-ray counts have been detected during the slew passages and therefore, the uncertainty of the X-ray flux is very high, although the detection itself is significant. This causes a rather strong bias for the flux level of sources at the detection limit because many more sources lie just below the detection threshold than just above. Statistical fluctuations shift many of these above the threshold while the number of sources that have nominal fluxes above threshold, but remain undetected due to fluctuations to lower count numbers, is considerably smaller. In effect, we inevitably overestimate the average X-ray flux of sources close to the detection threshold. Additionally, differences in the conversion factors might induce the systematically higher flux of the XMM then of *ROSAT*. Furthermore, we assume that the larger deviations are caused by intrinsic variability generated by flares. The RASS flux is less affected by flares because the high flare fluxes have a smaller weight for RASS sources due to their longer exposure times. Additionally some sources can only be detected during a flare in the XMMSL2 catalog because the quiescent emission is below the XMMSL2 detection limit. In Fig. 12 the detection limit of the XMMSL2 catalog is visible at ~5×10^{-13} erg cm^{-2} s^{-1}, while a detection limit of the RASS catalog is not noticeable because in the RASS catalog the exposure time, and hence, the detection limit is not constant over the sky but the exposure time varies between typically 400 s at the ecliptic equator and ~40 000 s at the poles.

6. Properties of the stellar sample

Next we discuss the X-ray properties of the crossmatched stellar XMMSL2 sources. We also address the nature of the identified stellar counterparts, where we restrict the discussion to unique identifications, i.e., sources with exactly one stellar counterpart. However, in Sect. 6.1 we use the stellar counterparts only to determine if the XMMSL2 source is stellar, and therefore, we also include sources with multiple counterparts.

6.1. Number-flux-density distribution

Figure 13 shows the derived number-flux-density distribution. At the X-ray bright end of the diagram, the number of sources decreases approximately linearly with increasing flux in the dou-

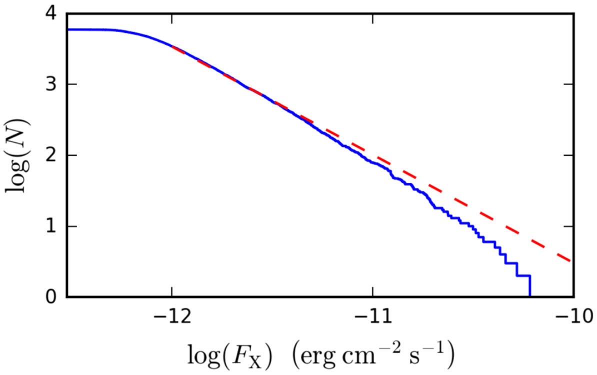


Fig. 13. Cumulative number-flux-density distribution of the stellar XMMSL2 sources. The red dashed line represents the best fit of the distribution.

ble logarithmic scale and we describe the distribution by the power law ansatz

$$N(S) = kF_{\mathrm{X}}^{-\alpha}, \tag{8}$$

where N is the number of sources with a flux brighter than F_{X}, α is the slope, and k is the normalization. Applying the method of (Crawford et al. 1970), the best value of α can be estimated by maximizing the likelihood function

$$L = M\ln(\alpha) - (\alpha+1)\cdot\sum_i \ln\left(\frac{F_{\mathrm{X},i}}{F_{\mathrm{X,min}}}\right) - M\ln\left(1-\left(\frac{F_{\mathrm{X,max}}}{F_{\mathrm{X,min}}}\right)^{-\alpha}\right), \tag{9}$$

where M is the total number of sources, $F_{\mathrm{X,max}}$ and $F_{\mathrm{X,min}}$ are the brightest and faintest flux, respectively, and $F_{\mathrm{X},i}$ is the flux of the ith source.

The slope depends on the lower flux limit imposed on the sample when fitting the distribution because there is no hard detection limit due to different exposure times and background levels. Using only sources brighter than $F_{\mathrm{X}} = 10^{-12}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$, we obtain a best value of $\alpha = 1.53 \pm 0.03$, which agrees to a spatially uniform distribution.

The best fit differs from the distribution of the brightest sources that show a slightly steeper slope. However, these sources have a smaller weight and higher uncertainty due to their small numbers. Most sources are found at faint fluxes (see the logarithmic scale in Fig. 13), where a slope of $\alpha = 1.53$ fits the distribution well.

6.2. X-ray luminosities

In Fig. 14 we present the distribution of the X-ray luminosities as a function optical color and compare the luminosities of our sample with those of the volume-limited NEXXUS sample (Schmitt & Liefke 2004). The dwarf stars in our sample are up to two orders of magnitude brighter in X-rays than the most active NEXXUS sources of the same spectral type. Even the faintest sources in our sample are more luminous than the Sun, which is found at $\sim 3\times10^{26}\,\mathrm{erg\,s^{-1}}$ and $\sim 5\times10^{27}\,\mathrm{erg\,s^{-1}}$ at solar minimum and maximum, respectively (Peres et al. 2000). Hence, we only see the high luminosity tail of the stellar luminosity distribution in our sample. In addition, we expect many sources to be detected during a flare, which further biases our sample to high X-ray luminosities.

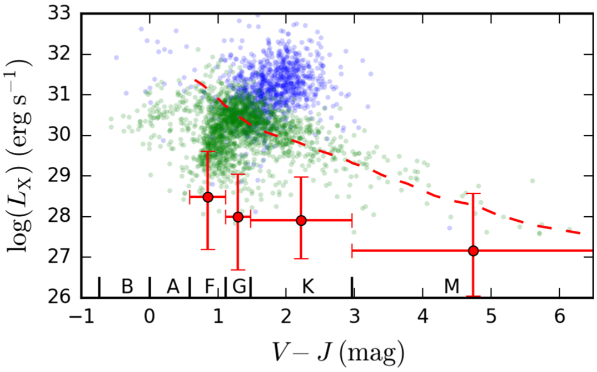


Fig. 14. X-ray luminosity distribution of the stellar XMMSL2 sources with known parallax. Main sequence stars and giants are shown as green and blue dots, respectively. The red crosses show the X-ray luminosities of the volume limited NEXXUS sample (Schmitt & Liefke 2004). The extension in x-direction corresponds to the width of the spectral type, while the extension in y-direction shows the luminosity range in which 90% of the NEXXUS dwarfs of the specific spectral type lie. The red dashed line represents the saturation limit at $L_{\mathrm{X}}/L_{\mathrm{bol}} = 10^{-3}$ for dwarfs.

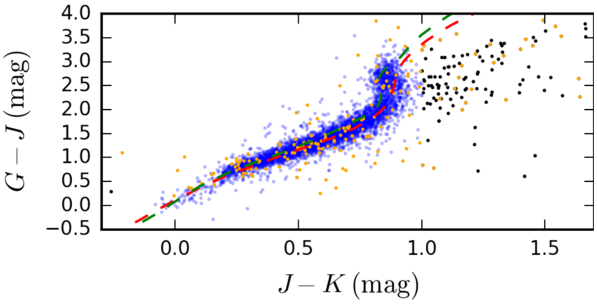


Fig. 15. Color-color diagram of the stellar XMMSL2 sources with *Gaia* and 2MASS counterparts. The sources represented as orange and black dots are flagged because of their known erroneous 2MASS photometry and their unusual colors, respectively. The red and green dashed curves represent various theoretical models of the main sequence; see text for details.

6.3. Color-color diagram

About 80% of the unique sources are covered by the 2MASS and *Gaia* catalogs, and we constructed color-color diagrams for those sources. Figure 15 shows that the sources are arranged in a well-defined streak. However, a few sources are located outside of the main distribution, many of which are known to have an erroneous 2MASS photometry. Many sources at the red end are classified as pre-main sequence stars by SIMBAD and their location in Fig. 15 is likely due to reddening. Since we do not perform any correction for extinction either in the optical or at X-ray wavelengths, we flag these sources in our catalog. Another possibility is an error in the optical magnitudes leading to incorrect colors, again, this motivates flagging these sources.

Figure 15 also shows two theoretical color-color relations for main sequence stars. We adopt color-color relations from Worthey & Lee (2011) and Pecaut & Mamajek (2013) to obtain the theoretical correlation between $G-J$ and $J-K$. The estimated theoretical color-color relation generally corresponds well with the observed distribution and differs only for the sources with

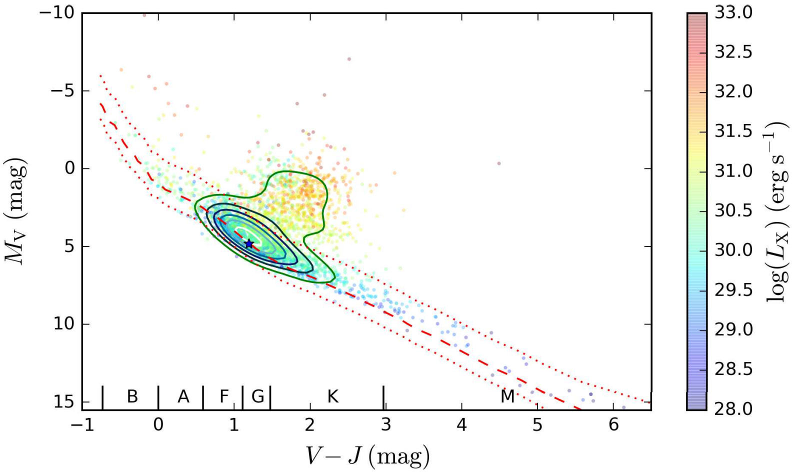


Fig. 16. Hertzsprung-Russell diagram of the stellar XMMSL2 sources with known parallax, the color scales with the logarithmic X-ray luminosity $\log(L_X)$. The red dashed line represents the theoretical main sequence following Pecaut & Mamajek (2013), and the dotted lines show the assumed width of the main sequence. The star marker represents the position of the Sun; we note that the Sun is less luminous than the stellar XMMSL2 sources. The density of the distribution is indicated by the colored lines. The ranges of the spectral types for dwarfs are given at the bottom of the figure.

$J - K \approx 0.8$. However, for these sources the predicted correlation between $G - J$ and $J - K$ substantially depends on the assumed model.

6.4. Hertzsprung-Russell diagram

About 57% of our sources with a unique stellar identification have trigonometric parallaxes, either from *Gaia* or the HIPPARCOS catalog (ESA 1997), or in the Lepine and BrightStar catalogs. Figure 16 shows the Hertzsprung-Russell diagram (HRD) of these sources. We point out that this sample of stellar counterparts with trigonometric parallaxes is neither complete in brightness nor in volume (many optically faint sources currently lack parallaxes), however, Fig. 16 does contain all basic features known from an optically selected HRD. The main sequence covers a wide range (-0.9 mag $< V - J < 6$ mag), hence the sample contains all spectral types from O-type stars down to dwarfs of spectral type M6V albeit sparsely populated at the limiting spectral types. Figure 16 contains only a small number of M-type dwarfs and some of them might actually be reddened K-type dwarfs, although late-type dwarfs show frequent and extreme flares. However, the fluxes at X-ray and optical wavelengths are low for these sources, and hence, the parallaxes are currently unknown. Furthermore, the complete sample also includes even later type stars, but these sources lack parallaxes and are mainly flagged because of the extreme color.

In the following, we assume sources to be main sequence stars if their absolute brightness M_V is in the range $M_{V,\mathrm{theo}} + 1.0 > M_V > M_{V,\mathrm{theo}} - 1.8$, where $M_{V,\mathrm{theo}}$ is the theoretical absolute brightness of a main sequence star adopted from Pecaut & Mamajek (2013). We find that 64% of the sources are dwarfs. However, not all of these sources need to be single stars, rather we expect many sources to be X-ray bright, active binaries. Also, the giant branch is clearly evident and we expect many of these sources to be RS CVn systems.

The HRD presented by Güdel (2004) contains about 2000 stars and shares many similarities with our HRD. However, the number of stars associated with the various categories differs strongly because our sample is drawn from a flux limited sample that is biased toward active systems, while Güdel (2004) congregated data from several studies. This also explains why we only find a few early-type stars and no separate population of pre-main sequence stars. Nevertheless, we expect that our complete sample contains some pre-main sequence stars, but these sources either lack parallaxes or are flagged because of reddening.

6.5. F_X/F_{bol}-ratio

Figure 17 shows the F_X/F_{bol} distribution of the stellar XMMSL2 sources as a function of the effective temperature (see Fig. 11 for the histogram of the distribution). We also include sources that are flagged in our catalog due to optical loading or their high F_X/F_{bol} values. The stars cover a wide range of activity, some sources have very high values of F_X/F_{bol}, and we find that about 14% of the sources have an X-ray activity of $\log(F_X/F_{\mathrm{bol}}) > -2.5$, which is substantially higher than the saturation limit found in previous studies (e.g., Pizzolato et al. 2003). For many of these sources the RASS flux is up to 10 times fainter (see Fig. 12) and we expect them to be detected by *XMM-Newton* during a flare. We flag sources with persistent (XMMSL+ RASS) high X-ray flux since these are probably nonstellar sources. Furthermore, we flag strong transients and sources lying above the red dashed line in Fig. 17; the high X-ray activity of these sources is unlikely caused by flares. Almost 6% of the sources lie above the threshold of $\log(F_X/F_{\mathrm{bol}}) > -2.2$ that we use to calculate the magnitude cutoff. Since this magnitude cutoff is particularly relevant at both ends of the spectral type distribution, some true stellar identifications might be missed, especially in the M dwarf regime. At very low masses, sources are flagged owing to the extreme red colors and are missing in this figure.

The X-ray activity of the stars strongly depends on the spectral type and several known features are visible in Fig. 17. First, the onset of convection at about 7000 K is evident. Second, two distinct so-called fingers at 7000 K and 5000 K contain sources down to relatively low activity levels. Most sources in the first finger are F dwarfs, while the second finger mainly consists of late G and early K giants. Since the X-ray flux has to be above the detection limit, the sources with a very low activity level must have a high (apparent) bolometric flux. It is also evident from the color coding that the low activity sources are relatively bright in the optical and are either nearby, inactive dwarfs or distant, X-ray bright giants. Third, there are few early-type stars that have a low X-ray flux compared to their bolometric flux.

We show the sources with more than 17 counts (10% of the samples) and, hence, with highly reliable X-ray fluxes as black

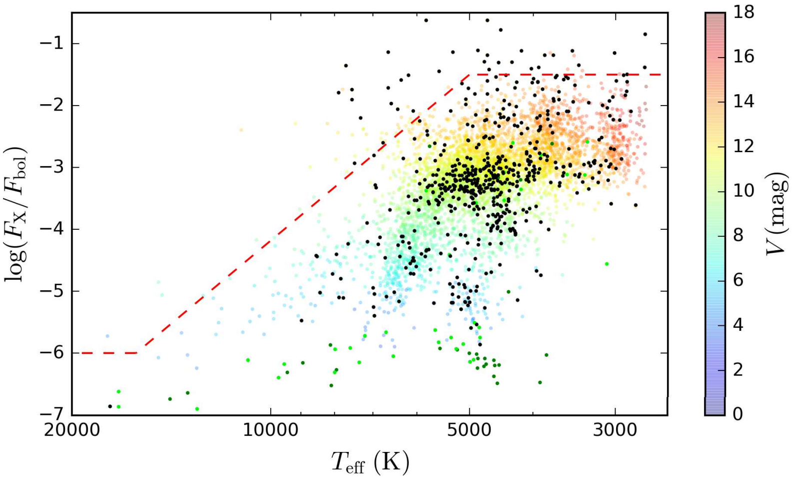


Fig. 17. F_X/F_{bol} distribution of the stellar XMMSL2 sources as a function of effective temperature. The colors scale with V band magnitude. Sources that may be affected by optical loading are shown as green dots, while those sources with a RASS identification are indicated in light green, and those sources without RASS counterpart are shown as dark green dots. The black dots represent highly reliable XMMSL2 sources with more than 17 cts (about 10 % of our sample). Sources lying above the red dashed line are flagged in our catalog as high X-ray activity during the slew.

dots in Fig. 17. Most of these sources have $\log(F_X/F_{bol}) < -3$, however, for some sources the contribution of the X-ray flux to the bolometric flux is extremely high. These sources could not be associated with a stellar counterpart brighter than the magnitude cutoff if the source had been detected at the XMMSL2 detection limit.

7. Conclusions and outlook

In this paper we present the first in-depth analysis of the stellar content of the *XMM-Newton* slew survey. In our approach the stellar XMMSL2 sources are identified by an automatic crossmatch of the XMMSL2 catalog with the *Gaia* DR1, 2MASS, and *Tycho*2 catalogs; we expect that in the future all necessary stellar data may actually be adopted from the new *Gaia* releases. We tested and verified the reliability of our automatic identification procedure by a comparison with the individually classified EMSS sources and by a crossmatch of our stellar identifications with the CSC.

With our procedure a total of 6815 stellar counterparts were found for the 23 252 XMMSL2 sources, after applying our cleaning procedures we identified 5920 XMMSL2 sources as high quality stellar X-ray sources corresponding to a stellar fraction of 25.5 further expect that 195 stellar classifications are spurious, i.e., they are erroneously attributed to a star. On the other hand, we expected 227 stellar identifications to be missed due to the chosen probability cutoff. Therefore our sample has a reliability of 96.7% and a completeness of 96.3%. We further expected a small fraction of true stellar identifications to be missed due to our magnitude cutoffs. This is a significant improvement to the stellar classifications presented by Saxton et al. (2008), who had no *Gaia* data at their disposal and identified only 4231 of our 5920 stellar XMMSL2 sources as stars. Saxton et al. (2008) further gave a stellar identification for 863 XMMSL2 sources that we did not identify as stars with our procedure, however, we show that most of the sources are less reliable stellar counterparts due to their high F_X/F_{bol} values.

The XMMSL2 catalog contains stars of all spectral types and luminosity classes, however, most of the stellar XMMSL2 sources are – not unexpectedly – late-type dwarfs with an outer convection zone. Only about 75% of the XMMSL2 sources have a RASS identification, although the RASS catalog is deeper by a factor of about 2.5. Hence, a substantial portion of the stellar XMMSL2 sources are previously unknown X-ray sources caught in an active or flaring state. With procedures as developed in this paper we expect to be able to perform an online identification of the stellar sources in the upcoming eROSITA all-sky survey.

Acknowledgements. This research has made use of data obtained from XMMSL2, the Second *XMM-Newton* Slew Survey Catalogue, produced by members of the XMM SOC, the EPIC consortium, and using work carried out in the context of the EXTraS project ("Exploring the X-ray Transient and variable Sky", funded from the EU's Seventh Framework Programme under grant agreement no. 607452). Additionally, we used data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC; `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. Furthermore, this publication made 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. Moreover, we used the VizieR catalog access tool and the SIMBAD database, operated at CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS 143, 23 and we have made use of data obtained from the *Chandra* Source Catalog, provided by the *Chandra* X-ray Center (CXC) as part of the *Chandra* Data Archive. SF acknowledge supports through the Integrationsamt Hamburg, the ZAV of Bundesagentur für Arbeit, and the Hamburg University, JR by DLR under grant 50 QR 1605 and PCS through the SFB 676 founded by DFG and by DLR under grant 50 OR 1706.

References

Agrawal, P. C., Rao, A. R., & Sreekantan, B. V. 1986, MNRAS, 219, 225
Ayres, T. R. 2009, ApJ, 696, 1931
Berghoefer, T. W., Schmitt, J. H. M. M., Danner, R., & Cassinelli, J. P. 1997, A&A, 322, 167
Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103
Carpenter, J. M. 2001, AJ, 121, 2851
Catura, R. C., Acton, L. W., & Johnson, H. M. 1975, ApJ, 196, L47
Crawford, D. F., Jauncey, D. L., & Murdoch, H. S. 1970, ApJ, 162, 405
ESA 1997, The Hipparcos and *TYCHO* catalogues. Astrometric and photometric star catalogues derived from the ESA Hipparcos Space Astrometry Mission (Noordwijk: ESA Publications Division) ESA SP, 1200
Dempsey, R. C., Linsky, J. L., Fleming, T. A., & Schmitt, J. H. M. M. 1993, ApJS, 86, 599
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37
Favata, F., Micela, G., Orlando, S., et al. 2008, A&A, 490, 1121
Fleming, T. A., Liebert, J., Gioia, I. M., & Maccacaro, T. 1988, ApJ, 331, 958

Gaia Collaboration (Brown, A. G. A., et al.) 2016a, A&A, 595, A2
Gaia Collaboration (Prusti, T., et al.) 2016b, A&A, 595, A1
Gioia, I. M., Maccacaro, T., Schild, R. E., et al. 1990, ApJS, 72, 567
Güdel, M. 2004, A&ARv, 12, 71
Haisch, B., Schmitt, J. H. M. M., & Rosso, C. 1991, ApJ, 383, L15
Hempelmann, A., Schmitt, J. H. M. M., Baliunas, S. L., & Donahue, R. A. 2003, A&A, 406, L39
Hoffleit, D., & Jaschek, C. 1991, The Bright Star Catalogue
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
Huensch, M., Schmitt, J. H. M. M., Schroeder, K.-P., & Reimers, D. 1996, A&A, 310, 801
Huensch, M., Schmitt, J. H. M. M., & Voges, W. 1998a, A&AS, 132, 155
Huensch, M., Schmitt, J. H. M. M., & Voges, W. 1998b, A&AS, 127, 251
Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Lépine, S., & Gaidos, E. 2011, AJ, 142, 138
Linsky, J. L., & Haisch, B. M. 1979, ApJ, 229, L27
Merloni, A., Predehl, P., Becker, W., et al. 2012, ArXiv e-prints [arXiv:1209.3114]
Pallavicini, R., Golub, L., Rosner, R., et al. 1981, ApJ, 248, 279
Pallavicini, R., Tagliaferri, G., & Stella, L. 1990, A&A, 228, 403
Pecaut, M. J., & Mamajek, E. E. 2013, ApJS, 208, 9
Peres, G., Orlando, S., Reale, F., Rosner, R., & Hudson, H. 2000, ApJ, 528, 537
Pevtsov, A. A., Fisher, G. H., Acton, L. W., et al. 2003, ApJ, 598, 1387
Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., & Ventura, P. 2003, A&A, 397, 147
Predehl, P. 2017, Astron. Nachr., 338, 159
Robrade, J., Schmitt, J. H. M. M., & Favata, F. 2012, A&A, 543, A84
Saxton, R. D., Read, A. M., Esquej, P., et al. 2008, A&A, 480, 611
Schmitt, J. H. M. M. 1997, A&A, 318, 215
Schmitt, J. H. M. M., & Liefke, C. 2004, A&A, 417, 651
Schmitt, J. H. M. M., Fleming, T. A., & Giampapa, M. S. 1995, ApJ, 450, 392
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Stelzer, B., Schmitt, J. H. M. M., Micela, G., & Liefke, C. 2006, A&A, 460, L35
Stocke, J. T., Morris, S. L., Gioia, I. M., et al. 1991, ApJS, 76, 813
Vaiana, G. S., Cassinelli, J. P., Fabbiano, G., et al. 1981, ApJ, 245, 163
Vilhu, O. 1984, A&A, 133, 117
Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389
Walter, F., Charles, P., & Bowyer, S. 1978, ApJ, 225, L119
Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P., & O'Dell, S. L. 2000, in X-Ray Optics, Instruments, and Missions III, eds. J. E. Truemper, & B. Aschenbach, Proc. SPIE, 4012, 2
Weisskopf, M. C., Brinkman, B., Canizares, C., et al. 2002, PASP, 114, 1
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Worthey, G., & Lee, H.-c. 2011, ApJS, 193, 1

Appendix A: Flowchart of the matching and cleaning procedure

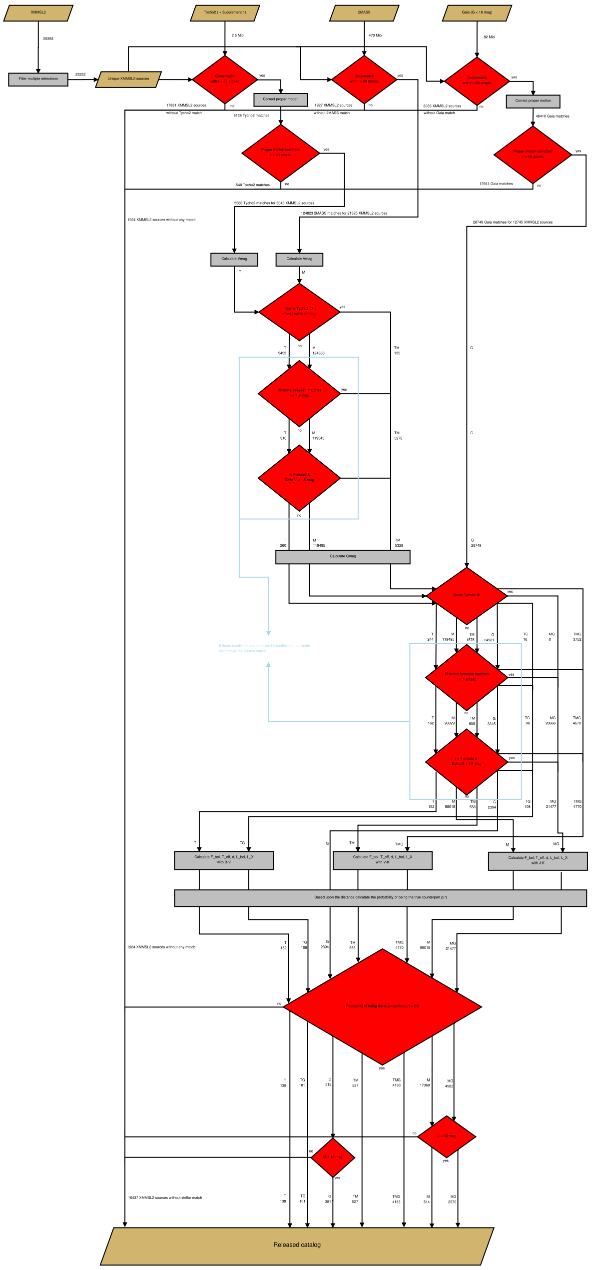


Fig. A.1. Flowchart of the matching procedure.

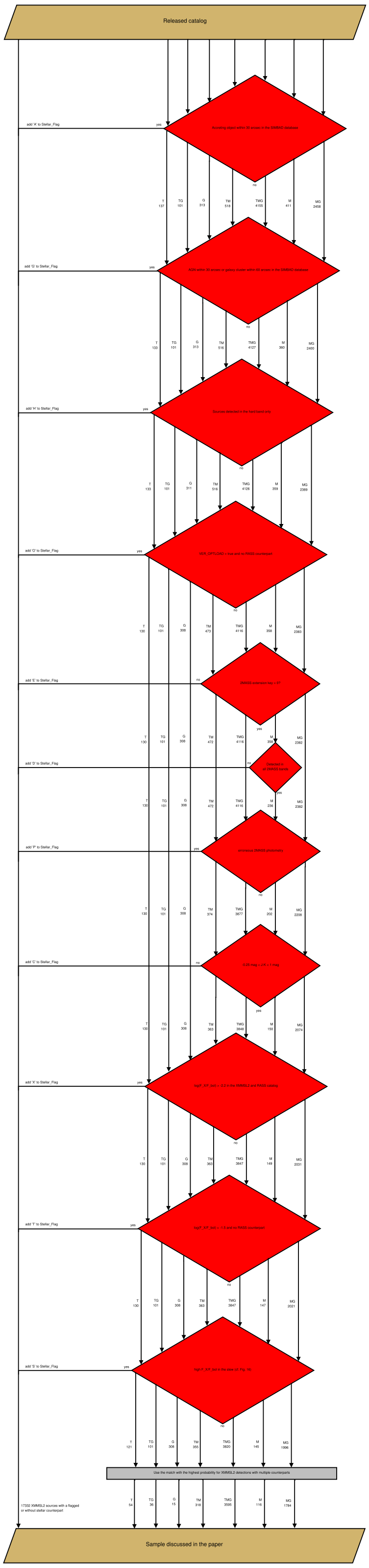


Fig. A.2. Flowchart of the cleaning procedure.

Appendix B: Column description

We adopt all columns of the XMMSL2 catalog, described online[3], and extend further 57 columns defining our stellar counterpart. In the following we describe these columns. For measurements given in multiple catalogs, we define an order of priority from which catalog the values are adopted.

XMMSLFX

We convert the count rates given in XMMSL2 catalog into a 0.1 – 2.4 keV X-ray flux (cf. Sect. 3.1). We use the count rate of the total band if available, otherwise we adopt the count rate of the soft band or the hard band.
Units: fW/m^2
Order of priority: total band, soft band, hard band

NumDet

Number of detections of the XMMSL2 source

NumMatch

Number of stellar counterparts of the XMMSL2 source

Priority

For XMMSL2 sources with multiple counterparts, the counterparts are sorted by the matching probability "Prob_r". This column gives the ranking of the counterparts so that the identification with the highest matching probability is set to "Priority=1".

Catalog

Catalog of the stellar counterpart
G: *Gaia* catalog
T: *Tycho*2 catalog
M: 2MASS catalog
L: Lepine catalog
B: BrightStar catalog
For counterparts given in multiple catalogs, multiple abbreviations are given, e.g., "TMG".

r

The distance r between the position given in the XMMSL2 catalog and the proper motion corrected position of the stellar counterpart
Units: arcsec
Order of priority: *Gaia* catalog, *Tycho*2 catalog, Lepine catalog, BrightStar catalog, 2MASS catalog

Prob_r

Matching probability as defined in Eq. (2) of the stellar counterpart
Units: %
Order of priority: Lepine catalog, BrightStar catalog, *Tycho*2 catalog, 2MASS catalog, *Gaia* catalog

GaiaID

Identifier of the *Gaia* catalog

*Tycho*2ID

Identifier of the *Tycho*2 catalog

HR

Identifier of the BrightStar catalog

PM

Identifier of the SUPERBLINK catalog given in the Lepine catalog

2MASSID

Identifier of the 2MASS catalog

MtchRA, MtchDE

Proper motion corrected position of the stellar counterpart
Units: degrees
Order of priority: *Gaia* catalog, *Tycho*2 catalog, Lepine catalog, BrightStar catalog, 2MASS catalog

e_MtchRA, e_MtchDE

Statistical error on the position of the stellar counterparts if it is available
Units: mas

pmRA, pmDE, e_pmRA, e_pmDE

Proper motion and their statistical errors of the stellar counterpart
Units: mas yr^{-1}
Order of priority: *Tycho*2 catalog, Lepine catalog, BrightStar catalog

BTmag, e_BTmag, VTmag, e_VTmag

Magnitude B_T and V_T and their statistical errors adopted from the *Tycho*2 catalog.
Units: mag

Bmag

The B magnitude is given in the Lepine catalog, but originates from the USNO catalog. It can also be estimated from the color $B - V$ given in the BrightStar catalog or from the *Tycho*2 colors (ESA 1997).
Units: mag
Order of priority: *Tycho*2 catalog, BrightStar catalog, Lepine catalog

Vmag

The V magnitude is directly measured in the BrightStar catalog. In the Lepine catalog the V magnitude is also given, but it is estimated from magnitudes of other photometric bands. We can estimate the V magnitude from the *Tycho*2 colors (ESA 1997) or

---

[3] `https://www.cosmos.esa.int/web/xmm-newton/xmmsl2-ug`

extrapolate it from the color $J - K$ of the 2MASS catalog.
Units: mag
Order of priority: *Tycho*2 catalog, BrightStar catalog, Lepine catalog, 2MASS catalog

Gmag

G magnitude adopted from the *Gaia* catalog.
Units: mag

Jmag, e_Jmag, Hmag, e_Hmag, Kmag, e_Kmag

Magnitudes in the J, H, and K bands and the statistical errors adopted from the 2MASS catalog.
Units: mag

Qflg

Quality of the 2MASS magnitude measurement adopted from the 2MASS catalog (ph_qual).

Rflg

This flag describes which method has been applied to determine the magnitude for every 2MASS photometric band (rd_flg). If the flag contains a “0”, the source is not detected in that band. We flag these sources if they do not have an entry in another catalog.

Bflg

Number of sources for which the 2MASS magnitude is estimated at the same time (bl_flg).

Cflg

This flag indicates if the magnitude or the position of a 2MASS source is contaminated by a nearby source (cc_flg).

Xflg

This flag indicates if a 2MASS point source lies within the boundaries of an extended 2MASS source (gal_contam).

XSCID

If a 2MASS point source is associated with an extended 2MASS source, this column gives the identifier of the extended source (ext_key).

plx, e_plx

Trigonometric parallax and their statistical errors if it is known
Units: arcsec
Order of priority: *Gaia* catalog, HIPPARCOS catalog, Lepine catalog, BrightStar catalog

CCDM

Identifier of the component of multiple star system adopted from the *Tycho*2 or the BrightStar catalog
Order of priority: *Tycho*2 catalog, BrightStar catalog

SpType

Spectral type adopted from the BrightStar or Lepine catalog
Order of priority: BrightStar catalog, Lepine catalog

mbol

We estimate the bolometric magnitude using Table 3 of Worthey & Lee (2011). For the calculation we use different colors in the following order of priority.
Units: mag
Order of priority: $V - K$, $B - V$, $J - K$

Fbol

Given the bolometric magnitude, we estimate the bolometric flux.
Units: fW/m^2

Teff

Effective temperature given in Table 3 of Worthey & Lee (2011)
Units: K
Order of priority: $V - K$, $B - V$, $J - K$

VMag, Mbol

Given the trigonometric parallax, we estimate the absolute magnitude.
Units: mag

Lbol, LX

Given the trigonometric parallax, we estimate the luminosity.
Units: W

StFlg

Table B.1. Stellar flag.

Sy.	Description	Definition	No.
A	Accretor warning	Accreting objects within 30 arcsec in the SIMBAD database	111
G	Extragalactic warning	AGN within 30 arcsec or a galaxy cluster within 60 arcsec in the SIMBAD database	116
O	Optical loading	“VER_OPTLOAD” is true and no RASS counterpart	59
E	Extended source	“ext_key” $\neq$ 0 in the 2MASS catalog	21
D	Missing 2MASS detection	2MASS counterpart only and “rd_flg” = 0 in one band	22
P	Erroneous 2MASS photometry	“ph_qual” = X, U, F or E in one band	467
C	Extreme color	$J-K < -0.25$ mag or $J-K > 1$ mag	311
H	Hard band detection only	No values given in “RATE_B8” and “RATE_B6”	3
X	Persistent high X-ray activity	$\log(F_X/F_{\rm bol} > -2.2$ in the XMMSL2 and RASS catalog	61
T	Transient X-ray source	$\log(F_{\rm X,XMMSL2}/F_{\rm bol} > -1.5$ and no RASS counterpart	22
S	High X-ray activity in the slew	Fractional contribution of the X-ray flux to the bolometric flux in the slew higher than usual for a source with the specific effective temperature (cf. Fig. 17)	123

2RXSID

Identifier of the closest RASS counterpart up to a distance of 60 arcsec

2RXSr

Distance between the XMMSL2 source and the closest RASS counterpart
Units: arcsec

2RXSFX

RASS flux of the closest RASS counterpart adopting the conversion factor defined by Schmitt et al. (1995)
Units: fW/m^2

SBDID

Identifier of the closest SIMBAD counterpart up to a distance of 30 arcsec

SBDr

Distance between the XMMSL2 source and the closest SIMBAD counterpart
Units: arcsec

SBDotype

Classification of the closest SIMBAD counterpart

Chapter 4

Identification with additional counterpart properties

As discussed in Sect. 2.4.1, the stellar content of the RASS have so far been investigated only for preselected subsamples of specific source types or by methods that are not specialized for stars and do not contain all RASS sources. In Freund et al., submitted (Sect. 4.1), the full stellar content of the RASS is identified for the first time. All stellar subtypes that were previously investigated individually are obtained in the coherent, flux-limited sample. The completeness and reliability of the identifications are directly estimated from the stellar probabilities that are provided for every RASS source.

In comparison to the XMMSL, the RASS sources have larger positional uncertainties of typically 13 arcsec. Furthermore, the sensitivity of RASS is higher with about 1.5×10^{-13} erg cm^{-2} s^{-1} in general and even lower values in regions with a high exposure near the Ecliptic poles. Therefore, counterparts with magnitudes down to $G = 19$ mag needs to be considered causing a much larger counterpart density. Hence, the stellar RASS sources cannot be identified only by the geometric properties. Instead, the possible associations are weighted by their X-ray to G band flux ratio and the counterpart distance in Freund et al., submitted (Sect. 4.1) by applying a Bayesian framework.

The article presented in Sect. 4.1 was submitted for publication in Astronomy & Astrophysics. I provided the major contribution to the identification and the examination of the results, however, the steps and ideas are the result of discussions with all coauthors.

To further validate the reliability of the identification procedure, I construct in Sect. 4.2 a test catalog, for which the type of every source is known. The estimation of the Bayes map applied for the weighting of the RASS sources contains some parameters of personal choice. In Sect. 4.3 I discuss in more detail how the Bayes map is constructed and how the details of the estimation affect the results.

4.1 Publication: The stellar content of the ROSAT all-sky survey

Astronomy & Astrophysics manuscript no. stellar_content_of_RASS
January 12, 2022

The stellar content of the ROSAT all-sky survey

S. Freund[1], S. Czesla[12], J. Robrade[1], P.C. Schneider[1], and J.H.M.M. Schmitt[1]

[1] Hamburger Sternwarte, Universität Hamburg, 21029 Hamburg, Germany
e-mail: `sebastian.freund@uni-hamburg.de`
[2] Thüringer Landessternwarte Tautenburg, Sternwarte 5, D-07778 Tautenburg, Germany

January 12, 2022

ABSTRACT

Aims. We present and apply a method to identify the stellar content of the *ROSAT* all-sky survey (RASS).
Methods. We performed a crossmatch between the RASS sources and stellar candidates selected from Gaia EDR3 and estimated stellar probabilities for every RASS source from the geometric properties of the match and additional properties namely the X-ray to G band flux ratio and the counterpart distances.
Results. A comparison with preliminary detections from the first *eROSITA* all-sky survey (eRASS1) show that the positional offsets of the RASS sources are larger than expected from the uncertainties given in the RASS catalog. From the RASS sources with reliable positional uncertainties, we identify 28 630 (24.9 %) sources as stellar, this is the largest sample of stellar X-ray sources to date. Directly from the stellar probabilities, we estimate the completeness and reliability of the sample to be about 93 % and confirm this value by comparison to the identification of randomly shifted RASS sources, preliminary stellar eRASS1 identifications, and results from a previous identification of RASS sources. Our stellar RASS sources contain sources of all spectral types and luminosity classes. According to their position in the color-magnitude diagram, many stellar RASS sources are young stars with ages of a few 10^7 yr or binaries. When plotting the X-ray to bolometric flux ratio as a function of the color, the onset of convection and the saturation limit are clearly visible. We note that later type stars reach continuously higher F_X/F_{bol} values probably due to more frequent flaring. The color distribution of the stellar RASS sources clearly differs from the unrelated background sources. We present the 3D distribution of the stellar RASS sources that shows a clear increase of the source density near known stellar clusters.

Key words. X-ray: stars – stars: activity – stars: coronae – stars: late-type – methods: statistical

1. Introduction

The X-ray observatory *ROSAT* (Trümper 1984), launched in June 1990, performed the *ROSAT* all-sky survey (RASS) between August 1990 and January 1991 with the positional-sensitive proportional counter (PSPC); the resulting catalogs have been published by Voges et al. (1999), and later using improved processing and analysis algorithms by Boller et al. (2016). Back in the old days, the X-ray properties of RASS detected stellar sources were investigated by selecting specific subsamples, e.g., RS CVn systems (Dempsey et al. 1993), OB-type stars (Berghoefer et al. 1997), or volume-limited or flux-limited samples of late-type and/or giant stars (Schmitt et al. 1995; Huensch et al. 1996; Schmitt 1997; Huensch et al. 1998b,c, 1999; Schmitt & Liefke 2004).

The RASS catalog has also been cross-correlated with optical and IR all-sky catalogs, for example identifications based on the *Tycho* and *Hipparcos* catalogs are presented by Guillout et al. (1999), while Haakonsen & Rutledge (2009) identified bright RASS sources with the *2MASS* Point Source Catalog, however, without providing a classification of the source types. Salvato et al. (2018) present an identification of high latitude RASS sources using the *AllWISE* catalog with a special emphasis on the identification of AGN; they also provide stellar identifications classified by a relation between X-ray and *AllWISE* fluxes. Because Salvato et al. (2018) focus on extragalactic objects, they exclude sources near the Galactic plane or the Large and Small Magellanic Clouds where confusion with Galactic foreground objects is an issue. Since we are specifically interested in these Galactic foreground objects and since a significant number of stellar sources lies, in fact, in the Galactic plane, the exclusion of this sky region is problematic from a stellar point of view. Furthermore, parallaxes and, thus, absolute fluxes have become available in the meantime thanks to *Gaia* (Gaia Collaboration et al. 2016, 2021), which provide important information for the identification of the stellar counterparts to ROSAT sources. In summary, the full stellar content of the RASS has so far never been identified and new data such as the *Gaia* survey put us in the position to accomplish this task.

The optical brightness of counterparts to stellar X-ray sources at a specific X-ray luminosity level can substantially vary due to different X-ray production mechanisms. The X-ray emission of early-type stars is thought to be generated through instabilities in their radiatively driven stellar winds and reach a typical fractional contribution of the X-ray to the total energy output of $L_X/L_{bol} \approx 10^{-7}$ (Pallavicini et al. 1981; Berghoefer et al. 1997); thus these OB-type stars are relatively X-ray faint when compared to their optical brightness. On the other hand, for late-type stars with a convective envelope, i.e., with stellar masses between 0.08 and 1.85 $M_\odot$ (spectral types M to mid A), the X-ray emission is thought to be produced by some magnetic field related heating mechanism in a hot corona. The X-ray activity strongly correlates with stellar rotation and age (Wilson 1963; Preibisch & Feigelson 2005), and hence, the X-ray luminosities substantially differ even for stars of the same mass.

The lowest levels of X-ray activity are observed at $L_X/L_{bol} \approx 10^{-8}$ (see Güdel 2004; Testa et al. 2015, and references therein) for old slowly rotating stars, while the X-ray emission saturates at $L_X/L_{bol} \approx 10^{-3}$ (Vilhu 1984; Wright et al. 2011) for very young and fast rotating stars. Also, evolved stars show X-ray emission with in some cases very high X-ray luminosities, especially when the rotation period is preserved by tidal interaction with a binary component as in RS CVn and related systems (Walter et al. 1978; Dempsey et al. 1993). Very little X-ray emission is found for red giants beyond the so-called dividing line (Linsky & Haisch 1979; Huensch et al. 1996). Furthermore, the observed X-ray luminosity of individual objects can increase over timescales of minutes to hours in some cases by orders of magnitude when observed during a flare. On timescales of years, the X-ray emission of late-type stars may show modulations related to activity cycles similar to the solar cycle (Hempelmann et al. 2003; Robrade et al. 2012).

Very importantly, as a consequence of the saturation limit, the stellar counterparts in X-ray surveys with a given limiting X-ray flux will also be limited in their optical flux, and therefore a complete census of the stellar RASS source requires a homogeneous counterpart catalog of sufficient depth. Furthermore, additional properties of the counterparts ought to be available to distinguish between stellar and non-stellar counterparts and between likely identifications and unrelated background sources within the relatively large error circles of the RASS positions. Such a catalog is provided by the *Gaia* mission (Gaia Collaboration et al. 2016) which contains, in its current version *Gaia* EDR3 (Gaia Collaboration et al. 2021), highly accurate positions, proper motions, magnitudes, colors, and parallaxes for more than 1.4 billion sources down to the 21st magnitude. Specifically, given a typical RASS sensitivity of 2×10^{-13} erg s^{-1} cm^{-2}, we expect most counterparts to be brighter than (*Gaia*) magnitude G $\approx$ 15 and be located within a distance of a few hundred parsec. Therefore we assume that all stellar counterparts of RASS X-ray sources are, first, contained in *Gaia* EDR3, and second, have a parallax measurement, which allows a placing of the source in the Hertzsprung-Russell diagram.

We structured our paper as follows: in Sect. 2 we present the input and matching catalogs namely the RASS catalog and *Gaia* EDR3, furthermore, we introduce the ***eROSITA*** all-sky survey (eRASS) that we use for verification. Next, we discuss the positional uncertainties of the RASS sources, define our sample of stellar candidates, and describe our identification procedure in Sect. 3. We present our results in Sect. 4 and compare our stellar identifications with eRASS1 and Salvato et al. (2018). In Sect. 5 we discuss the properties of the obtained sample of stellar RASS sources and we draw our conclusions in Sect. 6.

2. Input and matching catalogs

2.1. Second ROSAT all-sky survey (2RXS) source catalog

During its all-sky survey, the *ROSAT* satellite scanned the sky along great circles over the ecliptic poles resulting in an exposure time varying between about 400 s and 40000 s for most parts of the sky (Voges et al. 1999). From the reduction of these data, Boller et al. (2016) created the second *ROSAT* all-sky survey (2RXS) source catalog (hereafter: RASS catalog) and provide a detailed description of the content of the catalog.

For our work the following properties of the RASS catalog are of particular importance: the RASS catalog contains about 135 000 X-ray sources in the 0.1 – 2.4 keV energy band down to a detection likelihood of 6.5, yet, depending on detection likelihood, up to 30 % of the sources are expected to be spurious detections. Due to the observing strategy, the exposure time strongly varies with the ecliptic latitude, and hence, the RASS catalog has no uniform detection limit. We adopt the count rate to flux conversion of Schmitt et al. (1995) throughout this paper leading to a detection limit ranging from $\sim 3 \cdot 10^{-14}$ – $5 \cdot 10^{-13}$ erg s^{-1} cm^{-2} for 99 % of the RASS sources with a mean detection limit of approximately $1.5 \cdot 10^{-13}$ erg s^{-1} cm^{-2}. The statistical uncertainty of the positions of the RASS sources depends – partly – on the number of detected counts and we discuss the RASS positional uncertainty in Sect. 3.1.

2.2. Gaia EDR3

The current version of the *Gaia* catalog, *Gaia* EDR3, is based on 34 months of data collected between July 2014 and May 2017. In the following we provide a description of the properties and limitations of *Gaia* EDR3 that are most important in the context of this paper, and we refer to Gaia Collaboration et al. (2021) and Fabricius et al. (2021) for a detailed description.

Gaia EDR3 provides highly accurate positions with typically sub-milliarcsecond uncertainties at epoch J2016. More than 1.4 of the 1.8 billion *Gaia* EDR3 sources further contain proper motions and parallaxes, their typical uncertainties are smaller than 0.5 mas and strongly decrease with increasing brightness of the source. Most of the sources without proper motion and parallax are very faint, especially for sources fainter than about $G = 21$ mag only the position is provided, but also 1.5 % of the sources brighter than $G = 19$ mag miss parallaxes and proper motions in many cases due to problems with a close neighbor. Furthermore, some parallaxes are unreliable, specifically 1.6 % of the sources with `parallax_over_error` > 5 are expected to be spurious, the fraction strongly decreasing with source magnitude. The *Gaia* EDR3 catalog also contains broad band photometry in the G, BP, and RP band. For about 5.4 million sources, the G band magnitude is missing and nearly 300 million sources do not have a BP or RP magnitudes due to processing problems.

Gaia EDR3 is essentially complete between $G = 12$ and 17 mag, however, the completeness is reduced for sources brighter than $G = 7$ mag and 20 % of the stars brighter than 3 mag are missing. Furthermore, the catalog is incomplete in regions with source densities above 6×10^5 stars deg^{-2}. Binaries are resolved for separations above 2 arcsec and the completeness decreases rapidly for separations below 0.7 arcsec. Even for resolved binaries, the parallax or the magnitude in one band might be missing due to processing problems.

2.3. eROSITA all-sky survey

In July 2019 the extended ROentgen Survey with an Imaging Telescope Array (*eROSITA*) instrument was launched onboard the Russian-German Spectrum-Roentgen-Gamma (SRG) mission and started its four year lasting all-sky survey in December 2019; a detailed description of the *eROSITA* hardware, mission and in-orbit performance is presented by Predehl et al. (2021). The first ***eROSITA*** all-sky survey (eRASS1) was completed in summer 2020 and the preliminary data in the western half of the sky with Galactic longitudes $l > 180°$ are processed by the eSASS pipeline (currently version 946) (Brunner et al. 2021). We adopted preliminary stellar identifications of the eRASS1 catalog (Freund et al. (in prep.)) obtained by a Bayesian algorithm similar to the method described in Schneider et al. (2021) adopting a Bayes map as a function of the X-ray to G-band flux ratio and the $BP - RP$ color.

With a flux measurement obtained 30 years ago, the RASS survey provides a very important point of comparison to many sources detected by *eROSITA*, and, further, since the positional accuracy of the eRASS1 sources is much better (on average ~ 3 arcsec) than that of the RASS sources, the eRASS1 sources are ideally suited to check and verify the RASS positions and stellar identifications.

3. Matching procedure

We seek to identify the full stellar content of the RASS and, more specifically, the coronal X-ray emitters detected in the RASS. In this context we also include OB-type stars when we speak of coronal sources although the X-ray emission in these sources is produced by stellar winds, yet the X-ray properties are rather similar to proper coronal sources.

In our first paper, we identified the stellar content of the *XMM-Newton* slew survey (XMMSL) (Freund et al. 2018) applying only the angular separation as matching criterion. This is suitable for the XMMSL because the positions of the XMMSL sources are quite accurate and the density of the candidate stellar counterparts is quite small. Therefore, most XMMSL sources have only one plausible stellar counterpart. However, for the RASS sources we often find several stellar candidates in the search region because the RASS positions are less accurate and the RASS catalog is more sensitive than XMMSL and thus, the stellar counterparts are expected to be fainter. Therefore, additional source properties have to be considered to find the correct identification.

In Sect. 3.1 we discuss the positional uncertainties of the RASS sources and in Sect. 3.2 we define our candidate stellar counterparts and estimate their source density. We describe the estimation of the matching probabilities adopting only geometric properties in Sect. 3.3 and consider additional properties by applying a Bayes map in Sect. 3.4.

3.1. Positional uncertainties of the RASS sources

We tested the positional offsets of the RASS sources by cross-matching them with the eRASS1 catalog, the positional uncertainties of which are small compared to that of the RASS sources, and therefore, neglected in the following. In Fig. 1 we show the angular separations between the RASS sources and the nearest stellar eRASS1 counterparts scaling the separations with the positional uncertainties given in the RASS catalog. To assess the number of spurious associations, we also show in Fig. 1 the separations of the eRASS1 sources to randomly shifted RASS sources. The offsets are expected to be described by the Rayleigh distribution through

$$n(r) = \frac{r}{\sigma_{\mathrm{RASS}}^2} e^{-\frac{r^2}{2\sigma_{\mathrm{RASS}}^2}}, \tag{1}$$

where r is the angular separations between the RASS and eRASS1 sources and σ_{RASS} is the positional uncertainty given in the RASS catalog. However, as shown in Fig. 1, the Rayleigh distribution does not describe the offsets of the RASS positions well, and hence, we fitted the positional uncertainties in arcsec by applying the ansatz

$$\sigma_r = s \cdot \sqrt{\frac{(\mathtt{XERR} \cdot 45)^2 + (\mathtt{YERR} \cdot 45)^2}{2} + \sigma_{sys}^2}, \tag{2}$$

where `XERR` and `YERR` are the uncertainties in detector coordinates, converted to arcsec by a factor of 45, given in the RASS catalog, σ_{sys} is the systematic uncertainty, and s is a scaling factor.

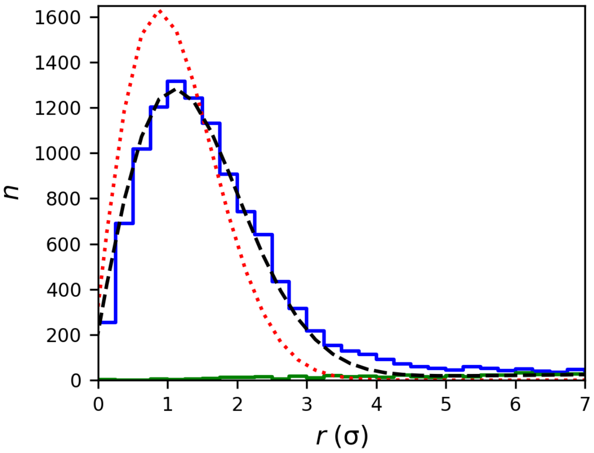


Fig. 1: Nearest neighbor distribution for the RASS and the stellar eRASS1 sources as function of the positional accuracy given in the RASS catalog. The blue and green solid histograms show the nearest neighbor distributions for the real and shifted RASS sources. The red and black dashed line show the Rayleigh distribution (Equation 1) and the best fit applying Equation 2, respectively.

We find that the nearest neighbor distribution can be fitted well by Equation 2 (see Fig. 1), however, a fraction of identifications still have a larger positional offset than expected from a Gaussian distribution and this fraction gradually increases with the positional uncertainty given in the RASS catalog. To restrict the fraction to about 5 %, we discuss in the following only the 115 000 RASS sources with the best positional accuracies, which we refer to as the main RASS catalog. For these sources, the nearest neighbor distribution is best fitted by a systematic uncertainty of $\sigma_{sys} = 3$ arcsec and a scaling factor of $s = 1.22$. Hence, the main RASS catalog only contains sources with a positional accuracy $\sigma_r < 20.4$ arcsec. We provide the identifications of the RASS sources with larger positional uncertainties in a supplementary catalog but we note that our matching procedure is less reliable for these sources.

3.2. Candidate counterparts and source densities

At the mean RASS detection limit, a high luminous stellar X-ray source with $L_X = 10^{31}$ erg s^{-1} can be detected up to a distance of ~750 pc. Since *Gaia* EDR3 provides accurate parallaxes out to much larger distances, we can differentiate between plausible stellar and extragalactic counterparts by only considering *Gaia* EDR3 sources with a parallax significance $> 3\sigma$, however, the exact value of the considered parallax value has a minor influence on our results.

In Fig. 2 we show the fraction of the sources that are filtered out by this parallax cutoff as a function of the source magnitude. Thus, the fraction of *Gaia* EDR3 sources with a low significance parallax that are probably extragalactic sources increases at $G \approx 17$ mag. We miss a small number of *Gaia* EDR3 counterparts that lack parallaxes due to processing problems, however, for a large magnitude range, this fraction is only about 1 % and slightly increases only for very bright and faint sources.

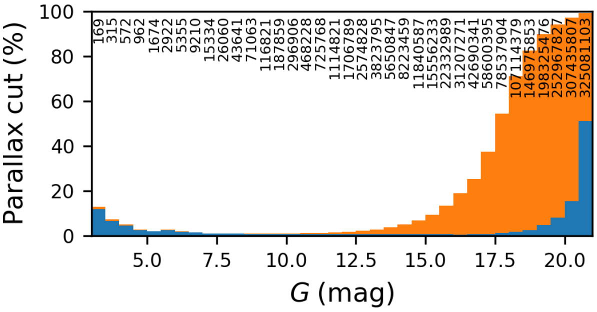


Fig. 2: Fraction of the *Gaia* EDR3 sources that are filtered out by the parallax cutoff as a function of the source magnitude. The blue bars indicate the fraction of sources without parallax in *Gaia* EDR3, while the orange bars represent the fraction of sources for which the accuracy of the parallax is less than 3σ. The values given at the top indicate the total number of *Gaia* EDR3 sources in the magnitude bin.

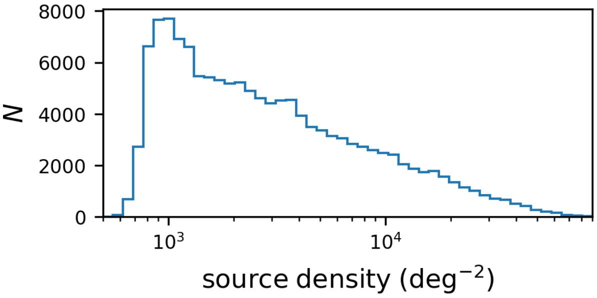


Fig. 3: Distribution of the density of eligible stellar candidates at the RASS positions

Due to the saturation limit, we further expect most of the RASS counterparts to be brighter than 15 mag and we excluded counterparts fainter than $G = 19$ mag to provide some margin for sources detected with a long exposure time or during a flare, again, the influence of the exact cutoff value is negligible. Furthermore, we restricted the sample to *Gaia* EDR3 sources with magnitude measurements in all three bands. We will henceforth refer to the so-selected sources as eligible stellar candidates. We then selected those candidate counterparts whose proper motion corrected angular separation from the RASS source is less than five times the positional accuracy of the RASS source.

Since *Gaia* EDR3 is incomplete at the bright end, we also crossmatched the RASS sources with the Tycho2 catalog. Again, we restricted the catalog to Tycho2 sources with magnitude measurements in the B_T and V_T bands and parallaxes from the Hipparcos catalog. From the Tycho2 B_T and V_T magnitudes, we estimated the brightness in *Gaia*'s G, BP, and RP band adopting the conversion provided in Busso et al. (2021). We assume *Gaia* EDR3 and Tycho2 matches to be associated to the same source if their angular separation is smaller than 2 arcsec and their extrapolated G band magnitude differs by less than 2 mag or the separation is smaller than 5 arcsec and the magnitude difference is less than 0.8 mag, however, the exact values do not influence the result significantly.

Since the density of the stellar candidates strongly varies between the Galactic plane and poles, the individual counterpart density at the position of the RASS sources also has to be considered in the estimation of the matching probability. Thus, we created an array of Hierarchical Equal Area isoLatitude Pixelation of a sphere (HEALpix, Górski et al. (2005)) pixels with a resolution of ∼ 27.5 arcmin and estimated the number of eligible stellar candidates in every pixel. Then, we adopted the number of sources in each pixel, where the RASS source is located, and divided this number by the pixel area to obtain the counterpart density at the location of every RASS source. For RASS sources located in HEALpix pixels with less than 40 plausible candidate counterparts, we increased the pixel size to obtain a source density that is less affected by statistical fluctuations. The resulting distribution of the source densities at the RASS positions is shown in Fig. 3. The density of eligible stellar candidates is about 10^3 deg^{-2} for most RASS sources but it increases to a few times 10^4 deg^{-2} for sources that are located close to the Galactic center.

3.3. *Determining the matching probabilities and the stellar fraction*

We deal with the problem of finding the correct identification to N_X X-ray sources, namely the RASS sources, in a catalog of N_O optical counterparts, namely our eligible stellar candidate *Gaia* EDR3 sources. Both the X-ray and optical catalog contain the celestial positions and the positional uncertainties of the sources and they provide additional properties, e.g. the count rate and the detection likelihood in the X-ray catalog and the magnitude and the parallax in the optical catalog. Since the positional accuracy of the *Gaia* sources is much higher than that of the RASS sources, we assume the errors of the optical positions to be negligible, and we further assume the X-ray positional uncertainties $\sigma_i, i = 1, N_X$ to be Rayleigh distributed.

The estimation of the matching probabilities in a Bayesian framework is discussed by Schneider et al. (2021). The approach compares the hypotheses $H_{ij}, j = 1, N_O$ that the i$^{\text{th}}$ X-ray source is associated with the j$^{\text{th}}$ counterparts and H_{i0} that the i$^{\text{th}}$ X-ray source is not associated with any of the counterparts. The prior probabilities that the j$^{\text{th}}$ optical counterpart is the identification of the i$^{\text{th}}$ X-ray source and that none of the optical counterparts are associated with the X-ray source is given by

$$P(H_{ij}) = \frac{p_s p_r}{N_O} = \frac{p_s p_r}{\eta\Omega} \tag{3}$$

$$P(H_{i0}) = 1 - p_s p_r, \tag{4}$$

where η and Ω are the counterpart density in the vicinity of the X-ray source and the area of the full sky, respectively. Since up to 30 % of the RASS sources are spurious detections that do not have an optical counterpart, we divided the catalog fraction used by Schneider et al. (2021) into a stellar fraction p_s and a fraction p_r of real (not spurious) X-ray detections, which is given in Table 1 of Boller et al. (2016). We estimated the likelihood of obtaining the data D_i considering only the geometric properties and given the hypotheses through

$$P(D_i|H_{ij}) = \frac{1}{2\pi\sigma_i^2} e^{-\frac{r_{ij}^2}{2\sigma_i^2}} \tag{5}$$

$$P(D_i|H_{i0}) = \frac{1}{4\pi}, \tag{6}$$

where r_{ij} is the angular separation between the X-ray source and the optical counterpart. In practice, counterparts with separations much larger than the positional uncertainty can be neglected.

The fraction of stellar sources in the X-ray catalog p_s is generally not known but can be estimated from the likelihood of the matching configuration. Specifically, the likelihood for a single X-ray source is derived by the summation of the hypotheses and the likelihood of the matching configuration for the full catalog is estimated by the product of the single sources through

$$\begin{aligned}\mathcal{L}_{\text{config}} &= \prod_{i=1}^{N_X} \sum_{j=0}^{N_O} P(H_{ij}) \times P(D_i|H_{ij}) \qquad (7)\\ &= \prod_{i=1}^{N_X} \left[(1 - p_s p_r) \times P(H_{i0}) + \frac{p_s p_r}{\eta\Omega} \times \sum_{j=1}^{N_O} P(D_i|H_{ij}) \right] \qquad (8)\end{aligned}$$

The best value of p_s is estimated by the maximum likelihood through

$$\frac{\partial \mathcal{L}_{\text{config}}}{\partial p_s} = \sum_{i=1}^{N_X} \frac{\frac{p_r}{\eta\Omega} \times \sum_{j=1}^{N_O} P(D_i|H_{ij}) - p_r \times P(H_{i0})}{\sum_{j=0}^{N_O} P(H_{ij}) \times P(D_i|H_{ij})} \overset{!}{=} 0. \qquad (9)$$

We apply the stellar fraction p_s resulting from Equation 9 to the prior probabilities and obtain the posterior probability that the j$^{\text{th}}$ counterpart is the correct identification through

$$p_{ij} = P(H_{ij}|D_i) = \frac{P(D_i|H_{ij}) \cdot P(H_{ij})}{\sum_{k=0}^{N_O} P(D_i|H_{ik}) \cdot P(H_{ik})}, \qquad (10)$$

and the probability that any of the counterparts is the correct identification, and hence, the X-ray source is stellar through

$$p_{\text{stellar}} = \frac{\sum_{k=1}^{N_O} P(D_i|H_{ik}) \cdot P(H_{ik})}{\sum_{k=0}^{N_O} P(D_i|H_{ik}) \cdot P(H_{ik})}. \qquad (11)$$

3.4. Considering additional properties

The estimation described in Sect. 3.3 only uses the geometric properties of the counterparts, i.e., the angular separation, the positional uncertainty, and the counterpart density. However, additional properties can be considered with a Bayes factor B_{ij} by expanding Equation 5 to

$$P(D_i|H_{ij}) = \frac{1}{2\pi\sigma_i^2} e^{-\frac{r_{ij}^2}{2\sigma_i^2}} \times B_{ij}. \qquad (12)$$

The construction of a Bayes factor, estimated by the fraction of the probability density functions (PDF) of the considered property for real stellar X-ray identifications and random associations, is described by Schneider et al. (2021).

To obtain a clean and unbiased sample of stellar X-ray sources, we selected 846 RASS sources with a geometric matching probability $p_{ij} > 0.9$, while 58 of these identifications are expected to be spurious. We inspected these sources individually and filtered out unlikely coronal X-ray emitters due to three non-exclusive categories; 38 and 99 sources are unlikely stellar identifications because of their high X-ray luminosity ($L_X > 10^{32}$ erg s^{-1}) and X-ray to G-band flux[1] ratio (applying an empirical relation[2]), respectively, and 67 sources in the sample are located more than 1.5 mag below the main sequence. These counterparts are probably the correct identification of the RASS source, but their X-ray emission is likely produced by a white dwarf or an accreting object and not by a corona. Excluding these unlikely coronal identifications leaves us with 736 sources that we refer to as training set in the following. We expect that a large fraction of the spurious identifications do not pass our filter criteria.

To be able to compare the properties of the true stellar identifications in the training set with spuriously identified background sources, we shifted all RASS sources randomly between 10 and 20 arcmin and by a random angle and select all eligible stellar candidates within $5\sigma_r$ to the shifted RASS sources as control set.

Since the true identifications in the training set and the random associations in the control set substantially differ in distance and X-ray to G-band flux, we created two dimensional Bayes maps of these properties. In Fig. 4 we show the distance distributions of the training and control set sources at different Galactic coordinates. The true identifications of the training set have small distances (typically < 400 pc), and hence, their distance distribution is not much affected by the Galactic structure. On the other hand, the background associations in the control set have much larger distances that increase to lower Galactic latitudes and towards the Galactic center. Therefore, we constructed different Bayes maps for each of the bins shown in Fig. 4. We applied the control set sources located in the individual bin but we adopted the whole training set for all bins because the sample size is small and the properties of the training set sources do not change significantly by Galactic coordinates. In App. A we show the Bayes maps at different Galactic coordinates. For example sources at 50 pc with $\log(F_X/F_G) = -4$ are upweighted, while counterparts at several kpc and $\log(F_X/F_G) = -1$ are downweighted. The dividing line between up- and downweighted regions slightly shifts to smaller distances and flux ratios with increasing Galactic latitude but the Bayes maps overall do not substantially change with Galactic coordinates.

4. Results

4.1. Stellar RASS identifications

We applied our matching procedure to the main RASS catalog and release the resulting catalog electronically at Centre de Données astronomiques de Strasbourg (CDS). In Appendix B we provide a detailed description of the released data.

We find 28 630 (24.9 %) RASS X-ray sources to be of stellar origin. However, the stellar fraction is not uniformly distributed over the sky, it increases towards the Galactic plane as shown in Fig. 5.

With our matching procedure, we can specifically compute the probability $p_{\text{stellar},i}$, that the i$^{\text{th}}$ X-ray source is of stellar origin. With this probability we can directly estimate of the number of missed and spurious identifications through the expressions

$$N_{\text{missed}} = \sum_{i}^{N_<} p_{\text{stellar},i} \qquad (13)$$

$$N_{\text{spurious}} = \sum_{i}^{N_>} (1 - p_{\text{stellar},i}), \qquad (14)$$

---

[1] We estimated the G-band flux through $F_G = 10^{-0.4\times G} \times W_{\text{eff}} \times ZP_\lambda$ adopting the effective bandwidth W_{eff} and the zero point ZP_λ from `http://svo2.cab.inta-csic.es/theory/fps/index.php?mode=browse&gname=GAIA&gname2=GAIA3&asttype=`

[2] $\log(F_X/F_G) > \begin{cases} (BP - RP) \times 1.7 - 3.58 & : BP - RP < 0.7 \text{ mag} \\ (BP - RP) \times 0.727 - 2.9 & : BP - RP > 0.7 \text{ mag} \end{cases}$

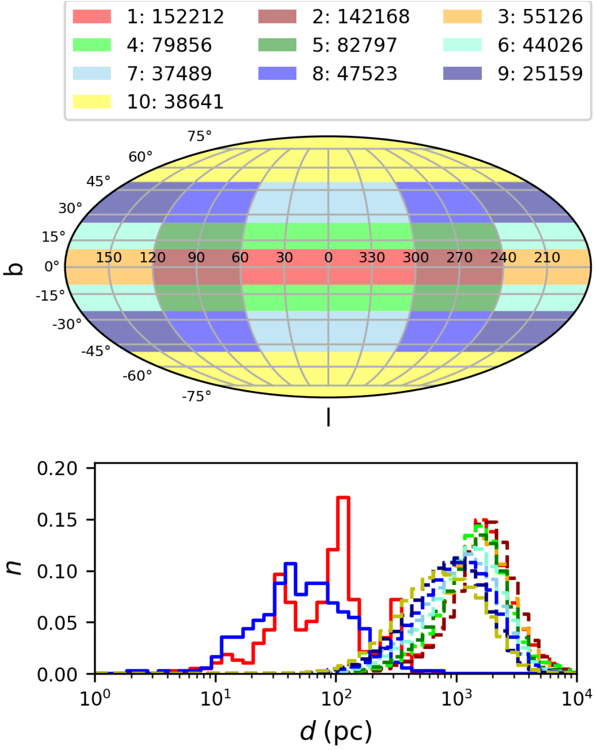


Fig. 4: Top: Bins in which the control set is divided in Galactic coordinates. The legend specifies the number of control set sources in the bins. Bottom: Distance distribution of the RASS counterparts at different Galactic coordinates. The solid red and blue histograms show the training set sources with $|b| < 25°$ and $|b| > 25°$, respectively, and the dashed lines show the distances of the control set sources, the colors refer to the bins shown in the top panel.

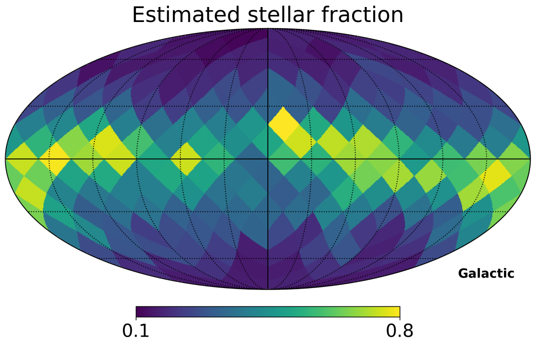


Fig. 5: Distribution of the estimated stellar fraction in Galactic coordinates

as well as the completeness and reliability through the expressions

$$\text{completeness} = \frac{N_> - N_{\text{spurious}}}{N_> - N_{\text{spurious}} + N_{\text{missed}}} \quad (15)$$

$$\text{reliabilty} = \frac{N_> - N_{\text{spurious}}}{N_>}, \quad (16)$$

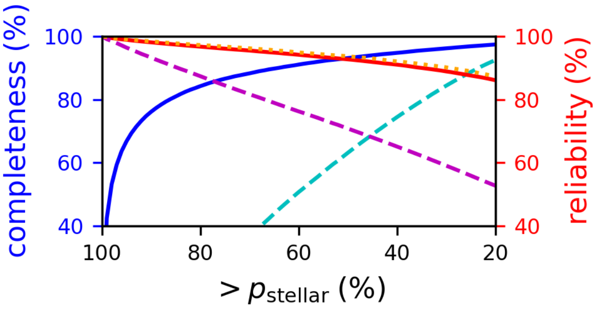


Fig. 6: Completeness and reliability of samples with different stellar probability cutoffs. For the dashed cyan and magenta line, only the geometric informations are used, while the solid blue and red curves show the completeness and reliability reached when the Bayes map is applied. The dotted orange line shows the reliability expected from shifted RASS sources

where $N_>$ and $N_<$ are the number of sources above and below a stellar probability cutoff. In Fig. 6 we show the resulting completeness and reliability of our stellar RASS identifications for samples with different stellar probability cutoffs. Naturally, a high cutoff value leads to a very reliable but incomplete sample, while relaxing the cutoff criterion increases the completeness at the cost of a lower reliability. At $p_{\text{stellar}} = 0.51$ the correct number of stellar sources is recovered and the completeness and reliability intersect at about 92.9 %. This is a significant improvement compared to considering only the geometric properties when a completeness and reliability of only about 68 % is reached. The number of spurious associations can be tested independently by applying the matching procedure to randomly shifted RASS sources counterparts which are spurious by definition. The so obtained estimate of the reliability fits the values derived from the stellar probability very well (see dotted orange line in Fig. 6) confirming the high accuracy of our estimated reliability.

Not all of the stellar RASS identifications can be unambiguously associated with one *Gaia* EDR3 stellar candidate, instead, 3621 RASS sources have two, 81 have three, and 4 have four plausible stellar counterparts with a matching probability $p_{ij} > 0.1$. Most of these sources are multiple star systems and the component responsible for the X-ray emission cannot be identified because of the large positional uncertainties of the RASS sources. In fact, the RASS source is likely a superposition of the emission from multiple components.

4.2. Comparison with the stellar eRASS1 sources

To investigate the reliability of our stellar RASS identifications, we consider X-ray sources detected during the first eROSITA all-sky survey; since the effective point response function of eROSITA and hence the positional accuracy of the eRASS1 sources is much better than that of the RASS sources, the eRASS1 sources are ideally suited to test our stellar RASS identifications. We specifically crossmatch the 52 842 RASS sources located at $l > 180°$ with the eRASS1 catalog and show in Fig. 7 the distribution of the angular separations. At small separations, the Gaussian peak of the true identifications is clearly visible, while the distribution approaches the linear slope of random associations at larger separations. As a compromise between spurious and missed identifications, we selected 32 363 RASS-

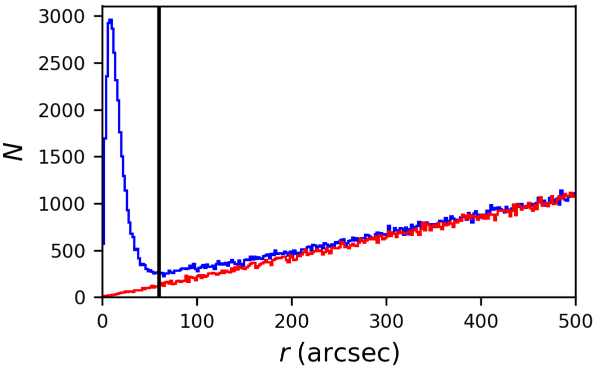


Fig. 7: Distribution of the angular separations between the RASS sources and the eRASS1 counterparts for the real (blue) and shifted (red) RASS sources. The vertical line shows our cutoff at 60 arcsec

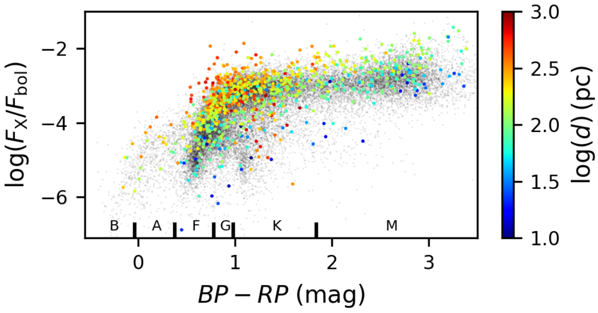


Fig. 8: F_X/F_{bol} vs. $BP - RP$ color for the sources identified as stellar by RASS and not by eRASS1. The color scales with the distance of the counterparts and the gray dots show the distributions for the whole stellar RASS sample as comparison. At the bottom we show the ranges of the spectral types as guidance.

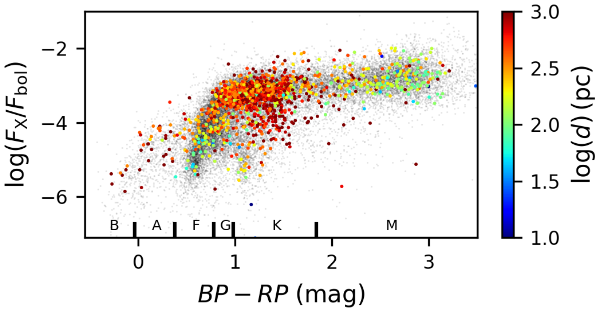


Fig. 9: F_X/F_{bol} vs. $BP - RP$ color for the sources identified as stellar by eRASS1 and not by RASS. The color scales with the distance of the counterparts and the gray dots show the distributions for the whole stellar RASS sample as comparison.

eRASS1 associations within an angular distance of 60 arcsec and identify 12 757 of these RASS sources with an eRASS1 counterpart as stellar; 12 692 sources have counterparts near or above the main sequence and thus, are likely coronal X-ray emitters (see Sect. 5.1).

The stellar identification based on RASS is confirmed by eROSITA for 11 811 (93 %) of the sources, hence, the predicted reliability of our stellar RASS sample is confirmed by eRASS1. In Fig. 8 we show the X-ray to bolometric flux[3] ratio as a function of the $BP - RP$ color for the stellar RASS identifications not confirmed by eRASS1 (cf Sect. 5.2). The spurious identifications not confirmed by eRASS1 tend to have slightly larger F_X/F_{bol} values, especially a few F- and G-type sources above the saturation limit are not confirmed. These RASS sources are identified because color is not considered by the weighting scheme. However, the sources not confirmed have overall similar properties as the correctly identified sources. On the other hand, 1291 (10.2 %) sources are classified as stellar by eRASS1 but not by RASS and we show the RASS X-ray bolometric flux ratios of these sources in Fig. 9, thus, the completeness is probably slightly lower than estimated from the stellar probabilities. This is caused by the fact that some RASS sources have larger positional offsets than expected from the given uncertainties (see Sect. 3.1), and furthermore, the eRASS1 catalog as well as the stellar identifications are preliminary and might be spurious in some cases. For example, some of these counterparts have large distances (> 1 kpc) that might be erroneously associated to the X-ray source because the preliminary identification procedure for eRASS1 currently does not consider the counterpart distances.

4.3. Comparison with NWAY

An identification of all RASS sources with Galactic latitudes in excess of 15° with the *AllWISE* catalog is presented by Salvato et al. (2018), who used NWAY, a Bayesian algorithm for identifying multi-wavelength counterparts adopting geometric parameters (e.g. angular separation, positional uncertainty, counterpart density) and *AllWISE* colors and magnitudes. NWAY provides a reliable identification (`p_any` > 0.5) for about 59 % to the RASS sources with about 5 % of them are expected to be random associations. The goal of Salvato et al. (2018) is to identify different source types in the RASS catalog but with a special emphasis on AGN, and therefore, they exclude RASS sources within the Galactic plane ($|b| < 15°$) and with separations smaller than 6° and 3° to the Large and Small Magellanic Clouds, respectively, to avoid source confusion in regions with high counterpart densities.

The catalog used by Salvato et al. (2018) has 90 850 sources in common with our main catalog, and we identify 18 739 of them as stellar. Salvato et al. (2018) also provide a relation to discriminate between stars and AGN, according to this criterion 19 679 of their best counterparts are stars. We compared the NWAY identifications with our results by considering *AllWISE* and *Gaia* EDR3 counterparts within 3.5 arcsec to be associated with the same source. We find that 17 284 (92 %) of the RASS sources are identified with the same counterpart by both methods, but note that 783 of them do not pass the stellar criterion from Salvato et al. (2018). The contradicting identifications are either associated by NWAY to an alternative stellar (499) or non-stellar (573) counterpart or the association is not reliable (`p_any` < 0.5) in NWAY (383). However, for 293 sources, our best identification is given by NWAY as second best counter-

---

[3] For the bolometric correction and the conversion between the $BP - RP$ and other photometric colors, we adopted throughout this paper the values given in a table based on Pecaut & Mamajek (2013) available at `http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt` (Version 2021.03.02)

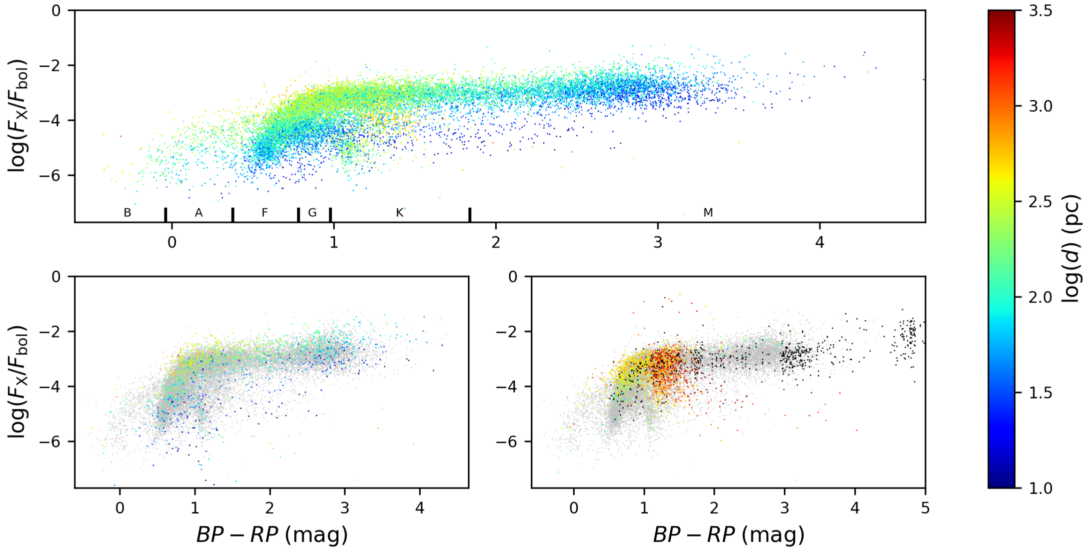


Fig. 10: Comparison of the X-ray to bolometric flux ratio as a function of the $BP - RP$ color for the stellar RASS sources with consistent and contradicting NWAY identification. The color scales with the distance of the counterparts, matches without parallaxes in *Gaia* EDR3 are shown in black. The top, left, and right panels show the distribution for the consistently identified sources, the sources identified by our method but not by NWAY, and the NWAY associations that we do not identify. The gray dots show the distribution of the top panel as comparison.

part. There are either two stellar counterparts (175) or a possible stellar and AGN identification (118), in both cases, the RASS source is likely a superposition of two X-ray emitters. On the other hand, 2703 (15.6 %) stellar counterparts are associated to a RASS source by NWAY but not by our procedure, 785 sources that we missed are identified by NWAY with a counterpart that is not in our list of stellar candidates. A crossmatch with the SIMBAD database (Wenger et al. 2000) reveals that some of these sources are indeed stellar sources that are not part of our stellar candidate list, mostly because of a missing parallax measurement in *Gaia* EDR3 but about 47 % of the sources are classified as extragalactic objects in SIMBAD.

In Fig. 10 we compare the X-ray to bolometric flux ratios of the RASS sources consistently identified by both methods with the flux ratios of sources that have contradicting identifications in NWAY and our method. For the counterpart only identified by NWAY without *Gaia* magnitude, we estimated the G band flux from *2MASS* magnitudes provided in the *AllWISE* catalog applying the conversion of Pecaut & Mamajek (2013). We note that many of the sources with different identifications have p_{stellar} values near the cutoff and hence, are not strictly excluded or confirmed as stellar sources by our method. Some of our stellar identifications not confirmed by NWAY are located at higher activity levels than the main distribution but overall, the identifications seem to be reasonable. A few of the sources only identified by NWAY as stellar are located far above the saturation limit and some have large distances of more than 1 kpc, we therefore argue that for many of these sources, the X-ray emission is unlikely to be produced by a stellar corona. Some of the NWAY identifications without Gaia counterpart have very red colors, most of them are classified as extragalactic objects in the SIMBAD database, and for these sources, the conversion between 2MASS and Gaia colors is not valid.

In summary, we conclude that despite NWAY being focused on identifications of extragalactic X-ray sources, the stellar identifications agree reasonably well with our results. The apparently higher number of identifications missed by our procedure is mainly caused by various incompletenesses that we expect to be removed in future *Gaia* releases. The stellar selection criterion by Salvato et al. (2018) produces somewhat different classifications so that about 4 % of our stellar identifications are missed as stellar by Salvato et al. (2018), and, further, some clearly non-coronal source types are considered stellar. Due to the fact that Salvato et al. (2018) considered only X-ray sources above Galactic latitudes of 15^{th} (and outside the Magellanic clouds), about a third of the stellar X-ray sky, i.e., nearly 10 000 stellar sources, is missed by NWAY.

5. Astrophysical properties of the stellar RASS sources

In the following we discuss the properties of the stellar RASS sources. To avoid ambiguous identifications, we restricted the analysis to 28 097 stellar RASS sources with $p_{\text{stellar}} > 0.51$ and $p_{ij} > 0.5$.

5.1. Color-magnitude diagram

In Fig. 11 we show the color-magnitude diagram (CMD) for the RASS sources identified as stars. Our X-ray selected sample contains dwarfs from late M-type to B-type. The number of counterparts with $BP - RP > 3$ mag (corresponding to a spectral type of M4V) is very small because these sources are very faint, and hence, unlikely to be detected in the RASS. We also find some O-type stars in the sample but they have rather large $BP - RP$ colors due to reddening, and hence, they are located in the regime of B-, A-, and F-type stars with absolute magnitudes typically larger than -3 mag. We note that no attempt to correct for extinction has been made neither in the optical nor in

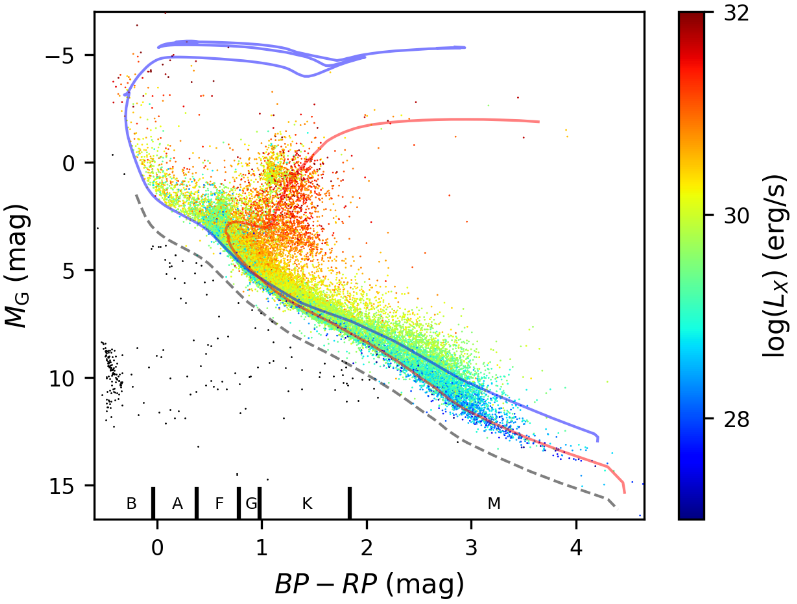


Fig. 11. Color-magnitude diagram of the RASS sources identified as stars. The color scales with the X-ray luminosities, and the blue and red solid lines show the PARSEC isochrones for stellar ages of 4×10^7 and 4×10^9 years (Bressan et al. 2012), respectively. Sources located below the dashed line (black dots) are more than 1.5 mag fainter than the main sequence.

the X-ray because most stellar RASS sources are located within a few hundred parsec. However, pre-main sequence and O-type stars located in molecular clouds and sources at large distances, e.g. giants and early-type stars, can be substantially affected by reddening.

Many of the late-type stellar RASS sources are located at positions in the CMD compatible with young stars with ages of a few 10^7 yr. Some of the stellar RASS sources might also be binaries unresolved in *Gaia* EDR3 that are shifted to lower magnitudes in Fig. 11. The sources above the main sequence have steadily increasing X-ray luminosities.

Also the giant branch is clearly visible in Fig. 11. Many of the X-ray detected giants have large luminosities of more than 1×10^{31} erg s^{-1}, we expect these sources to be active binaries like RS CVn systems. We find some RASS counterparts in the red giant branch beyond the so-called dividing line and we inspected some of these sources individually. A few sources are known spectroscopic binaries, hence, the X-ray emission might be produced by a late-type companion. However, especially many of the objects with large X-ray luminosities are known symbiotic binaries where the X-ray emission is produced by an accretion process (cf. Huensch et al. 1998a).

We also find some white dwarfs, although we did not train our matching procedure for these sources. Specifically, 223 sources are located more than 1.5 mag below the main sequence and we excluded these sources from the following analysis. We expect that most of these counterparts are the correct identification of the RASS source but the X-ray emission is unlikely to be produced by a coronal source but by a compact object, for example through accretion.

5.2. X-ray to bolometric flux ratio

We show the fraction of the X-ray to bolometric fluxes as a function of the $BP-RP$ color for the stellar RASS sources in Fig. 12. At spectral types of early F, where stars begin to develop an outer convection zone, the number of sources and the $F_X/F_{\rm bol}$ rapidly increases. We also find some A-type counterparts, although especially early A-type stars are generally not expected to produce X-ray emission Schmitt et al. (1985), here, a late-type companion is likely responsible for the X-ray emission in most cases.

While some stellar RASS sources have X-ray to bolometric flux ratios below 10^{-7}, the $F_X/F_{\rm bol}$ distribution peaks around the saturation limit at 10^{-3} and the distribution rapidly decreases for larger flux ratios. The width of the decrease is defined by the intrinsic scatter of the saturation limit and by flares during the RASS detection. Very few sources are found with $F_X/F_{\rm bol} > 10^{-2}$, they are likely detected during a large flare, and furthermore, the fraction of spurious identifications is larger for these sources than for the whole sample.

In Fig. 13 we compare the distributions of the X-ray to bolometric flux ratios for different colors. The bluest sources have very low flux ratios because they have no or very shallow convection zones and their X-ray emission is often produced by a late-type companion. Also very few of the F- and G-type stars reach flux ratios of 10^{-3} because only the youngest stars are saturated. Some of the sources with $1 < BP - RP < 1.5$ mag have low flux ratios, most of them are giants, but very few later type stars at small $F_X/F_{\rm bol}$ values are detected in the RASS due to the detection limit. Instead, the distribution of the stellar RASS sources later than $BP - RP = 1$ mag show a strong peak around $F_X/F_{\rm bol} \approx 10^{-3}$. The peak of the distributions and the highest flux ratios are shifted to higher values for later-type stars. The reason is most likely that the late-type sources flare more frequently and the flares have a larger impact on the $F_X/F_{\rm bol}$ values due to their small X-ray and bolometric fluxes in quiescence.

5.3. Color distribution of detected and not-detected stellar sources

To compare the properties of the stellar RASS sources with stellar sources not associated to a RASS detection, we selected stellar candidates located close to a randomly shifted RASS sources. In Fig. 14 we compare the $BP-RP$ color distributions of the stellar RASS sources with the counterparts to shifted RASS sources. The counterparts to the real RASS sources have significantly

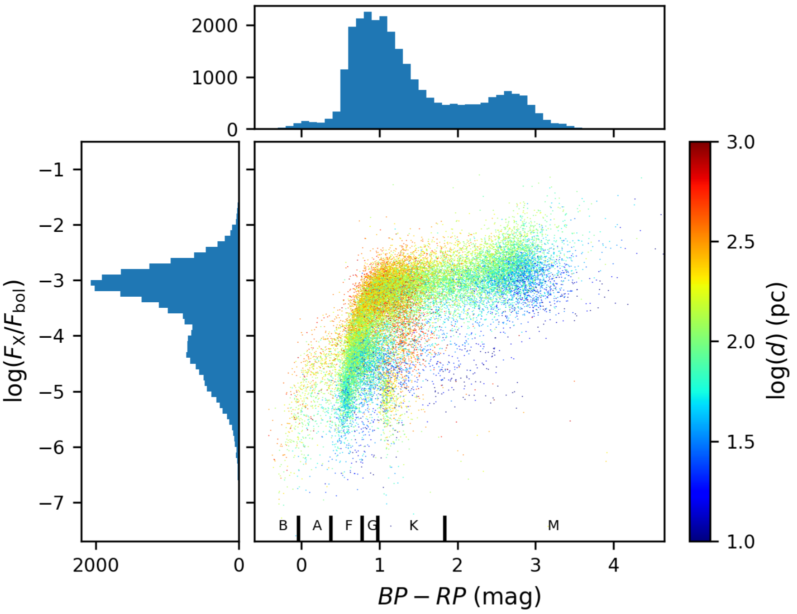


Fig. 12: $F_X/F_{\rm bol}$ vs. $BP-RP$ color for the RASS sources identified as stars. The color scales with the distances of the counterparts. The histograms show the distribution of the X-ray to bolometric flux ratios and the $BP-RP$ color, respectively.

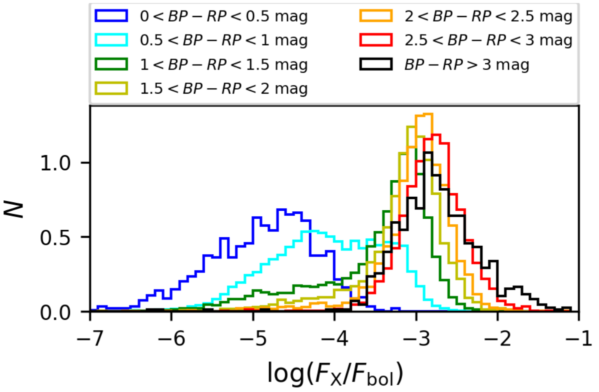


Fig. 13: Comparison of the $F_X/F_{\rm bol}$ distribution for the stellar RASS sources with different $BP-RP$ colors

different properties than the stellar candidates not associated to a real RASS source. First, both distributions show a sharp increase for F-type stars but while this increase is located at early F-type for the real RASS counterparts, the number of uncorrelated sources increases at late F-type. Furthermore, the true stellar RASS sources show a broad second peak around spectral type M3 that is not visible for the counterparts for the shifted sources.

The stellar candidates identified with a RASS source are generally brighter and have smaller distances than the candidates not associated to a RASS source. Therefore, we show in Fig. 14 the $BP-RP$ color distributions of the uncorrelated sources brighter than $G = 13$ mag and closer than 800 pc. The bright sources show a sharp increase at early F-type as the real RASS sources and the close stellar candidates have a bimodal distribution as the stellar RASS sources, although the distribution of the red stars peaks at slightly earlier spectral types. We note that a combination of bright and nearby stellar candidates does not fit the color distribution of the stellar RASS sources, and therefore, we conclude that the properties of the stellar candidates detected in the RASS significantly differ from the undetected sources.

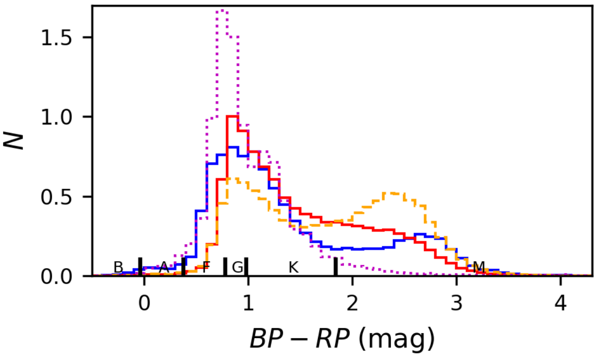


Fig. 14: Comparison between the normed $BP-RP$ color distributions of the stellar candidates identified with a RASS source (blue) and the stellar candidates not associated with a RASS source (red). The dashed orange and dotted magenta histograms show the color distributions for the unassociated candidates within 800 pc and brighter than $G = 13$ mag, respectively.

5.4. Three dimensional distribution

We estimated Galactic Cartesian coordinates for the stellar RASS sources from their *Gaia* positions and parallaxes; an interactive representation of the three dimensional distribution of the

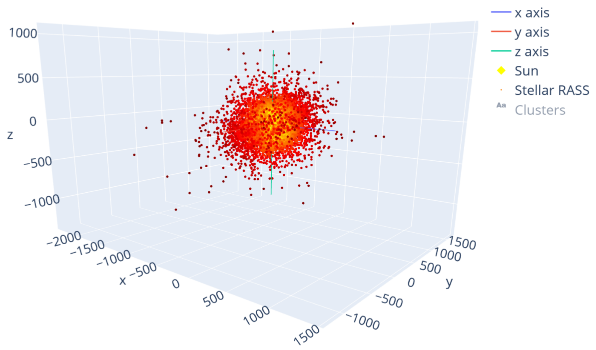


Fig. 15: Three dimensional distribution of the stellar RASS sources in Galactic Cartesian coordinates, where the x, y, and z axes are directed towards the Galactic center, the Galactic rotation, and to the Galactic north pole. The color scales with the distances of the stellar RASS sources to the sun.

stellar RASS sources is available online[4]. In Fig. 15 we show a two dimensional projection of the distribution for the whole sample. Most of the stellar RASS sources are located within 600 pc to the Sun, however, there are some outliers with distances of more than 1 kpc, but note that most of them have quite low stellar probabilities (typically $p_{\text{stellar}} < 0.7$), and hence, the fraction of spurious identifications is rather large. Due to their low distances from the Sun, the distribution of the stellar RASS sources is almost spherical with only a slight flattening caused by the thickness of the Galactic plane. The source density strongly increases with decreasing distance to the Sun due to the RASS detection limit.

When zooming in, several open clusters are visible. In Fig. 16 we show a 2D projection of a region closer to the sun, here, the Hyades, Pleiades, α Persei, and o Velorum clusters are visible and the density of stellar RASS sources is clearly enhanced in these regions. In the background at the top of Fig. 16, the Orion Nebula is located, in that region the RASS sources are clustered in many small structures.

6. Conclusion

In this paper we have presented the first identification attempt of the whole stellar content detected in the *ROSAT* all-sky survey. This is to our knowledge the largest sample of stellar X-ray sources presented so far. We adopted an automatic procedure to identify the RASS sources with stellar candidate counterparts selected from bright ($G < 19$ mag) *Gaia* EDR3 sources with a parallax significance $> 3\sigma$. Our procedure considers geometric informations of the match (angular separation, positional accuracy of the RASS sources, and counterpart density) as well as additional properties, namely the counterpart distances and the X-ray to G-band flux ratios.

A crossmatch with preliminary eRASS1 sources shows that the positional offsets of some RASS sources are larger than expected for Gaussian distributed positional uncertainties. This deviation increases with the positional uncertainty given in the RASS catalog, and hence, we discuss in this paper only the identifications of the 115 000 RASS sources with the highest positional accuracies. We also publish a supplementary catalog with the stellar counterparts of the remaining RASS sources but for those sources our identifications are less reliable.

Of the highly accurate RASS sources, we identify 28 630 (24.9 %) as stars and provide for each RASS source a stellar probability. From these probabilities, we estimate that our identifications are to about 93 % complete and reliable. We confirmed this value by comparisons with identifications to randomly shifted RASS sources, preliminary stellar identifications of eRASS1 sources, and the results from Salvato et al. (2018). Thus, we might miss a small number of stellar RASS sources due to incompletenesses in *Gaia* EDR3, a problem we expect to be reduced or remedied by future *Gaia* releases, and furthermore, some counterparts to the RASS sources have larger positional offsets than expected from Gaussian distributed uncertainties. The identifications from Salvato et al. (2018) generally agree well with our results, but the stellar classification by Salvato et al. (2018) that is solely based on X-ray and AllWISE fluxes seems to be less reliable compared to our selection of stellar candidates. In contrast to Salvato et al. (2018), we also provide stellar identifications in the crowded regions of the Galactic plane and near the Large and Small Magellanic Clouds where the association is more difficult, and hence, we identify about 35 % more stellar RASS X-ray sources.

The CMD of the stellar RASS sources shows all basic features known from an optically selected CMD. The main sequence contains dwarfs of spectral types from late M- to B-type, the number of counterparts with spectral types later than M4 are very rare due to their faintness. We also find many sources in the giant branch but for most of the identified red giants, the X-ray emission is probably produced by a late-type companion or a compact accreting object. Many of the stellar RASS sources are young stars with ages of a few 10^7 years according to their position in the CMD or their absolute magnitudes are increased by an unresolved companion. The X-ray to bolometric flux ratios strongly increase for early F-type stars due to the onset of convection and near $F_X/F_{\text{bol}} = 10^{-3}$ the saturation limit is clearly visible. The peak value of the F_X/F_{bol} distribution and the maximum values increase to later spectral types because these sources have stronger and more frequent flares. The color distribution of the stellar candidates detected in the RASS clearly differ from the not detected sources. Thanks to the parallaxes of *Gaia* EDR3, we can for the first time access the three dimensional distribution of the stellar RASS sources. Overall the stellar RASS sources are nearly spherically distributed, on smaller scales, stellar clusters with increased source densities are clearly visible.

The currently ongoing eROSITA all-sky survey is expected to detect more than 20 times as many stellar X-ray sources as identified in the RASS. With an identification algorithm similar to one presented in this paper, we are confident to be able to identify the stellar eRASS sources and make optimal use of the unprecedented potential of eROSITA to stellar X-ray science.

Acknowledgements. We thank Mara Salvato for useful comments and suggestions. SF gratefully acknowledge supports through the Integrationsamt Hildesheim, the ZAV of Bundesagentur für Arbeit, and the Hamburg University, SC by DFG under grant CZ 222/5-1, JR by DLR under grant 50 QR 2105, and PCS by DLR under grant 50 OR 1901 and 50 OR 2102. SF thanks Gabriele Uth and Maria Theresa Lehmann for their support. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work is based on data from eROSITA, the soft X-ray instrument aboard SRG, a joint Russian-German science mission supported by the Russian Space Agency (Roskosmos), in the interests of the Russian Academy of Sciences represented by its Space Research Institute (IKI), and the Deutsches

---

[4] Link will be added (TBD)

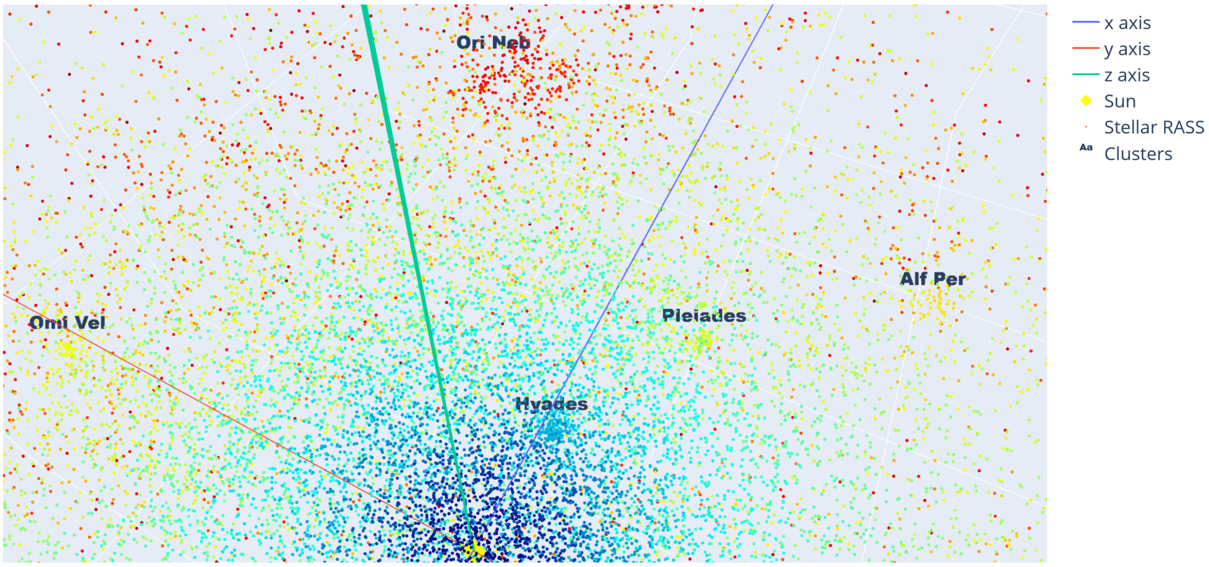


Fig. 16: Close-up of the three dimensional distribution of the stellar RASS sources in Galactic Cartesian coordinates. From the point of view of the plot, we are looking from the Galactic north pole on the Galactic plane (green z axis), along the blue x axis in the direction of the Galactic anti-center and the red y axis is directed against the Galactic rotation to the upper left. The color scales with the distances of the stellar RASS sources to the sun, whose position is shown by the yellow diamond.

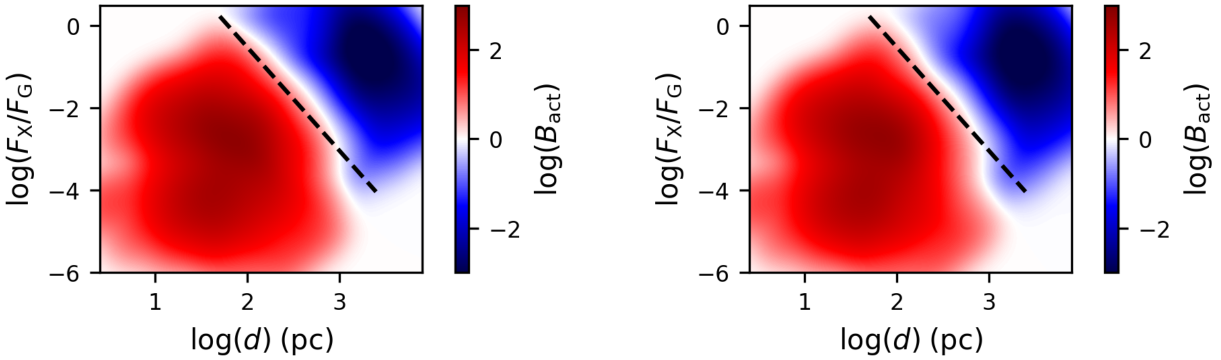
Zentrum für Luft- und Raumfahrt (DLR). The SRG spacecraft was built by Lavochkin Association (NPOL) and its subcontractors, and is operated by NPOL with support from IKI and the Max Planck Institute for Extraterrestrial Physics (MPE). The development and construction of the eROSITA X-ray instrument was led by MPE, with contributions from the Dr. Karl Remeis Observatory Bamberg & ECAP (FAU Erlangen-Nürnberg), the University of Hamburg Observatory, the Leibniz Institute for Astrophysics Potsdam (AIP), and the Institute for Astronomyand Astrophysics of the University of Tübingen, with the support of DLR and the Max Planck Society. The Argelander Institute for Astronomy of the University of Bonn and the Ludwig Maximilians Universität Munich also participated in the science preparation for eROSITA. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

References

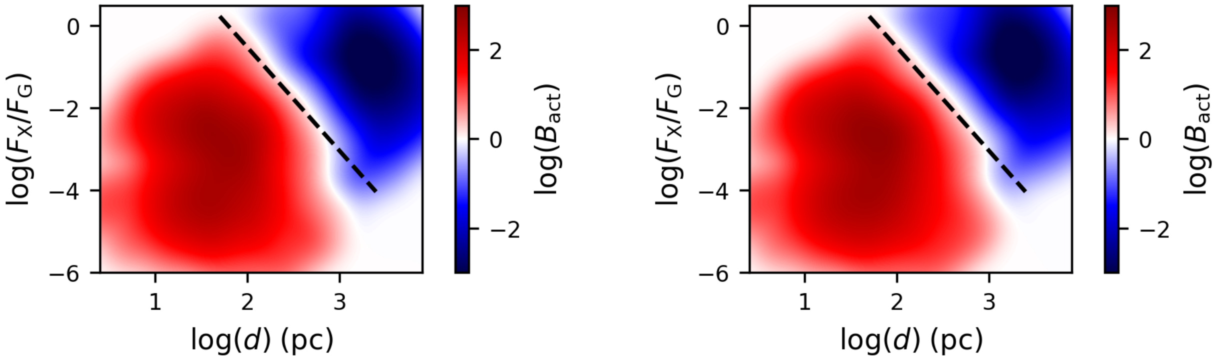
Berghoefer, T. W., Schmitt, J. H. M. M., Danner, R., & Cassinelli, J. P. 1997, A&A, 322, 167
Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Brunner, H., Liu, T., Lamer, G., et al. 2021, arXiv e-prints, arXiv:2106.14517
Busso, G., Cacciari, C., Bellazzini, M., et al. 2021, Gaia EDR3 documentation Chapter 5: Photometric data, Gaia EDR3 documentation
Dempsey, R. C., Linsky, J. L., Fleming, T. A., & Schmitt, J. H. M. M. 1993, ApJS, 86, 599
Fabricius, C., Luri, X., Arenou, F., et al. 2021, A&A, 649, A5
Freund, S., Robrade, J., Schneider, P. C., & Schmitt, J. H. M. M. 2018, A&A, 614, A125
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Güdel, M. 2004, A&A Rev., 12, 71
Guillout, P., Schmitt, J. H. M. M., Egret, D., et al. 1999, A&A, 351, 1003
Haakonsen, C. B. & Rutledge, R. E. 2009, ApJS, 184, 138
Hempelmann, A., Schmitt, J. H. M. M., Baliunas, S. L., & Donahue, R. A. 2003, A&A, 406, L39
Huensch, M., Schmitt, J. H. M. M., Schroder, K.-P., & Zickgraf, F.-J. 1998a, A&A, 330, 225
Huensch, M., Schmitt, J. H. M. M., Schroeder, K.-P., & Reimers, D. 1996, A&A, 310, 801
Huensch, M., Schmitt, J. H. M. M., Sterzik, M. F., & Voges, W. 1999, A&AS, 135, 319
Huensch, M., Schmitt, J. H. M. M., & Voges, W. 1998b, A&AS, 127, 251
Huensch, M., Schmitt, J. H. M. M., & Voges, W. 1998c, A&AS, 132, 155
Linsky, J. L. & Haisch, B. M. 1979, ApJ, 229, L27
Pallavicini, R., Golub, L., Rosner, R., et al. 1981, ApJ, 248, 279
Pecaut, M. J. & Mamajek, E. E. 2013, ApJS, 208, 9
Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
Preibisch, T. & Feigelson, E. D. 2005, ApJS, 160, 390
Robrade, J., Schmitt, J. H. M. M., & Favata, F. 2012, A&A, 543, A84
Salvato, M., Buchner, J., Budavári, T., et al. 2018, MNRAS, 473, 4937
Schmitt, J. H. M. M. 1997, A&A, 318, 215
Schmitt, J. H. M. M., Fleming, T. A., & Giampapa, M. S. 1995, ApJ, 450, 392
Schmitt, J. H. M. M., Golub, L., Harnden, F. R., J., et al. 1985, ApJ, 290, 307
Schmitt, J. H. M. M. & Liefke, C. 2004, A&A, 417, 651
Schneider, P. C., Freund, S., Czesla, S., et al. 2021, arXiv e-prints, arXiv:2106.14521
Testa, P., Saar, S. H., & Drake, J. J. 2015, Philosophical Transactions of the Royal Society of London Series A, 373, 20140259
Trümper, J. 1984, Physica Scripta Volume T, 7, 209
Vilhu, O. 1984, A&A, 133, 117
Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389
Walter, F., Charles, P., & Bowyer, S. 1978, ApJ, 225, L119
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wilson, O. C. 1963, ApJ, 138, 832
Wright, N. J., Drake, J. J., Mamajek, E. E., & Henry, G. W. 2011, ApJ, 743, 48

Appendix A: Bayes maps

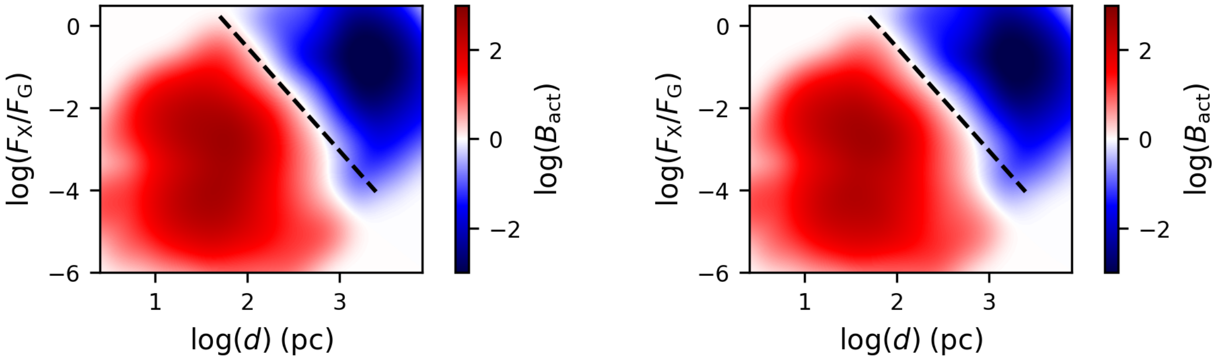
In Fig. A.1 we show the Bayes maps at the different Galactic coordinates. We applied a Gaussian kernel density estimator to estimate the PDFs of the training and control set and adopted a rather large bandwidth compared to Scott's rule to better smooth statistical fluctuations in our training set. Furthermore, we added a small constant to the PDFs so that the Bayes factor approaches unity in regions sparsely populated by training and control set sources. For example optically very bright sources with a low X-ray to G band flux ratio at large distances or sources that are very near to the sun but very faint in the optical are very rare in the training and control set, and hence, we do not weight these sources. Due to sparse populations, also the weighting for very bright and near sources as well as faint and large distant sources is reduced which might seem unphysically. However, since the number of such candidates is extremely small, this has little to no influence on our stellar identifications. The same is true for the other details of the estimation of the Bayes map.



(a) Bin 1 (b) Bin 2

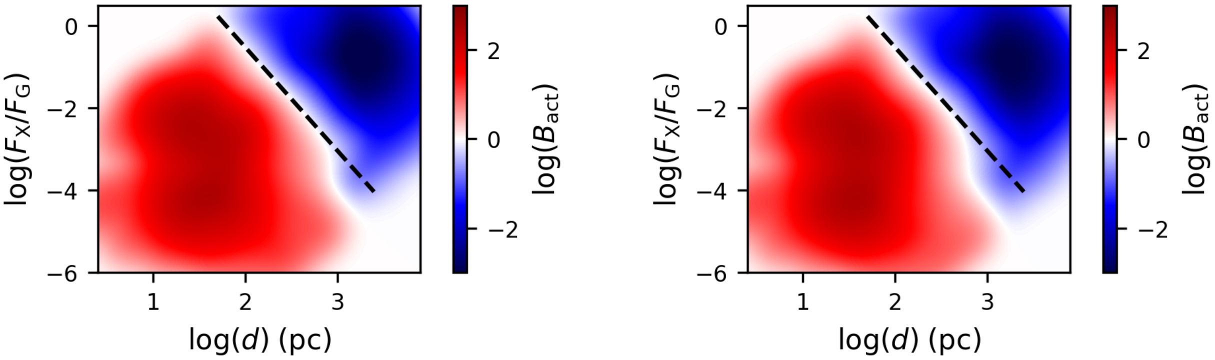


(c) Bin 3 (d) Bin 4



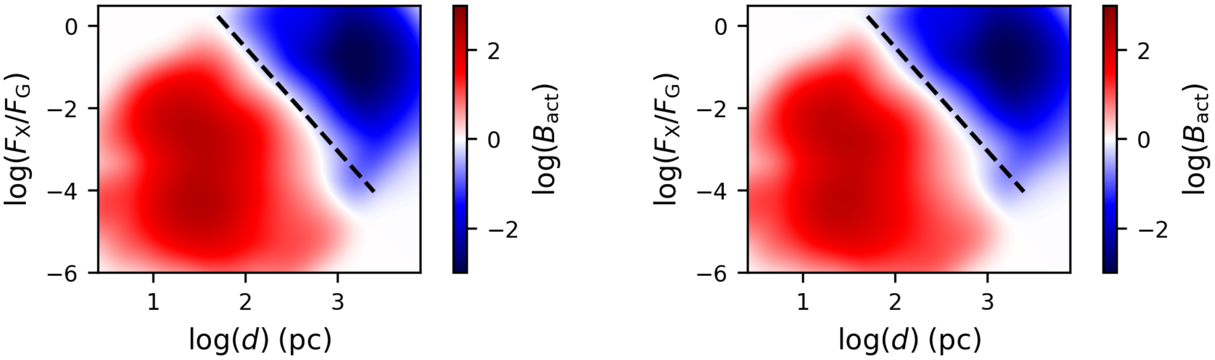
(e) Bin 5 (f) Bin 6

Fig. A.1: Bayes maps for the bins shown in Fig. 4. The dashed line is plotted at the same position in every panel as reference.



(a) Bin 7

(b) Bin 8



(c) Bin 9

(d) Bin 10

Fig. A.2: Bayes maps for the bins shown in Fig. 4 continued

Appendix B: Catalog release

In Table B.1 we provide the first 11 entries of our main matching catalog adopted through out this paper. The full catalog is available electronically at Centre de Données astronomiques de Strasbourg (CDS) containing the stellar counterparts with a matching probability $p_{ij} > 0.1$ to all RASS sources with a high quality positional accuracy and a stellar probability $p_{\mathrm{stellar}} > 0.2$. A supplement catalog with the counterpart of the RASS sources with a low positional uncertainty is also available at CDS but for these sources, our identification procedure is less reliable.

The catalog contains the names of the RASS sources and the stellar matches (*Gaia* source ID or Tycho2 ID if the source is not available in *Gaia* EDR3), the positional uncertainties of the RASS sources estimated by Equation 2, and the matching separations between the RASS sources and the stellar identifications. Furthermore, we provide the stellar probabilities (p_stellar) and the matching probabilities (p_ij) of the individual counterparts. Table B.1 also lists the proper motion corrected coordinates, the RASS X-ray fluxes, the G-band magnitudes, the $BP-RP$ colors, and the parallaxes of the counterparts.

RASS sources with multiple plausible counterparts have multiple entries in the catalog. For the discussion of the properties of the stellar RASS sources, we adopted the counterparts with $p_{\mathrm{stellar}} > 0.51$ and $p_{ij} > 0.5$ of the main catalog.

Table B.1: Basic properties of the stellar counterparts to the RASS sources

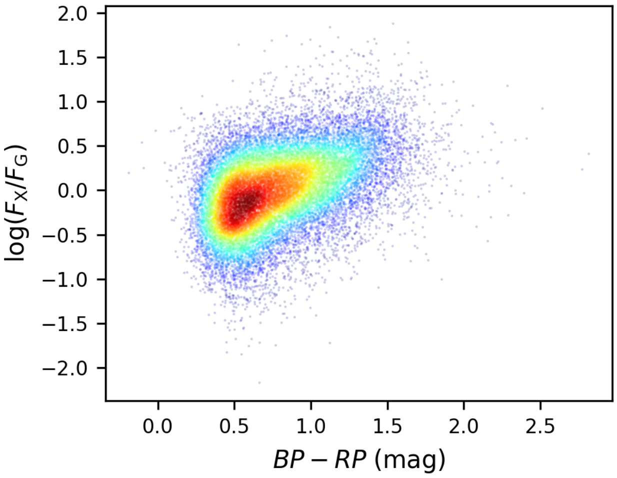
RASS_NAME	match_ID	sigma_r arcsec	sep arcsec	p_stellar	p_ij	match_RA deg	match_Dec deg	Fx erg/s/cm2	G mag	BP_RP mag	plx mas	subdwarf
2RXS J111220.4+872612	1151502176829452672	10.34	17.11	0.9571	0.9571	168.185944	87.435271	4.55e-14	13.04	1.21	4.01	False
2RXS J121506.0+874154	1728674141358771840	6.02	10.24	0.9996	0.9996	183.836210	87.699983	1.01e-12	6.19	0.46	20.67	False
2RXS J095440.9+872818	1151646212853601536	17.40	24.46	0.8186	0.8176	148.516947	87.472113	3.34e-13	14.61	2.73	11.61	False
2RXS J144219.2+871438	1728461317139734016	11.00	25.33	0.5936	0.5908	220.697854	87.239816	1.82e-13	13.49	1.32	4.14	False
2RXS J123746.9+875804	1728789727518860416	13.05	17.83	0.9917	0.9917	189.430636	87.972964	2.37e-13	8.99	0.65	8.61	False
2RXS J154734.6+872205	1728088578401606400	17.53	3.78	0.9757	0.9757	236.912721	87.367554	7.35e-14	12.47	1.39	8.28	False
2RXS J105207.8+884320	1152608349951908992	13.55	11.37	0.9905	0.9902	162.909618	88.724061	5.99e-14	12.02	1.31	6.89	False
2RXS J160911.9+875830	1728963450356277632	8.15	10.48	0.9951	0.9951	242.240334	87.977098	1.28e-13	8.91	0.76	10.39	False
2RXS J162322.1+883607	1729043474187041536	11.79	21.85	0.7308	0.7308	245.654162	88.606030	9.56e-14	8.15	1.05	3.34	False
2RXS J064236.2+880446	1152331242957967104	6.84	2.53	0.9978	0.7838	100.640436	88.078970	1.28e-12	10.69	0.87	5.78	False
2RXS J064236.2+880446	1152331238662558464	6.84	4.62	0.9978	0.2141	100.678522	88.078695	1.28e-12	12.22	1.10	5.55	False

4.2 Applying the identification procedure to a test catalog

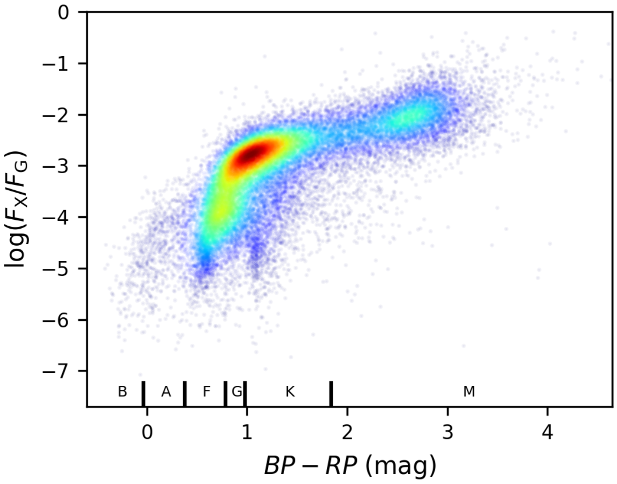
To validate the identification procedure, I created a test catalog, for which the source type of each object is known. Specifically, the test catalog contains simulated stellar, AGN, and spurious sources with randomized positions, positional uncertainties, X-ray fluxes, and existence likelihoods from sources of the main RASS catalog (cf. Freund et al., submitted; Sect. 4.1). Since the correct identification of every test source is known by construction, the reliability of the identification procedure can be checked directly. Although the test catalog was finally not included in Freund et al, submitted (Sect. 4.1), the well controlled sample was very important to understand the influence of the details of the identification procedure and to validate the reliability of the identifications.

4.2.1 Source selection

As the basis for the test catalog, I first selected 76 million *Gaia* EDR3 sources within 1 kpc as stellar sources and more than 1.6 million known extragalactic objects in *Gaia* EDR3 (called *Gaia*-AGN sample in the following) that were used to define the reference frame (*Gaia*-CRF3; Klioner et al., in prep.). To obtain realistic X-ray sources, I derived the properties of the test sources from known RASS identification. Specifically, I adopted the sample of stellar RASS sources described in Freund et al., submitted (see Sect. 4.1) for the stellar test sources. For the extragalactic objects, I crossmatched the *Gaia*-AGN sample with the RASS catalog and obtained more than 22 000 AGN within 1.5σ of the positional uncertainty of the RASS source expecting ∼ 2 % of them to be spurious identification. In Fig. 4.1 I compare the X-ray to G band flux ratios as a function of the $BP - RP$ color for the AGN and stellar sources that I used to construct the X-ray properties of the test sources. Both samples cover quite different areas in the F_X/F_G vs. BP-RP plane with the AGN being much fainter in the G band at the same X-ray flux.



(a) Extragalactic RASS sources



(b) Stellar RASS sources

Figure 4.1: X-ray to G band flux ratio as a function of the BP-RP color for a sample of extragalactic and stellar RASS sources. The color scales with the density of the sources.

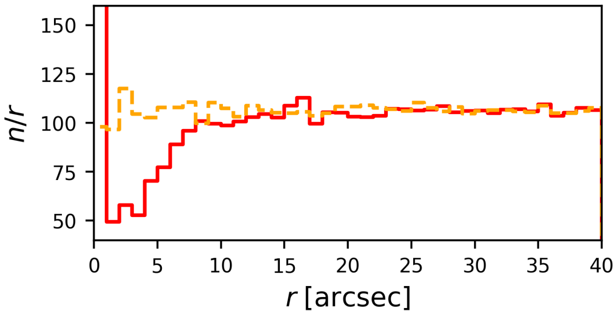


Figure 4.2: Distribution of the angular separation between the extragalactic sources and *Gaia* EDR3 counterparts. The solid red and dashed orange lines show the separation of the real and randomly shifted *Gaia*-AGN sources, respectively.

However, the positions of the *Gaia*-AGN sample are biased. In Fig. 4.2 I show the distribution of the angular separations between the *Gaia*-AGN sources and all *Gaia* EDR3 counterparts within 40 arcsec. There is a strong peak at small separations caused by the true associations but, although sources with separations larger than 2 arcsec are resolved in *Gaia* EDR3, the number of counterparts is reduced up to about 10 arcsec. Hence, the *Gaia*-AGN sources tend to be located in regions with a reduced local source density. The reason is that a large fraction of the extragalactic objects is adopted from a sample of Secrest et al. (2015) identified with the AllWISE catalog. The angular resolution of *WISE* is much lower than for *Gaia* (Wright et al., 2010), and therefore, AGN with a visually near stellar source in the foreground cannot be identified by Secrest et al. (2015). However, a common source of error of the identification procedure is the spurious identification of a stellar counterpart with an X-ray emission that is caused by a background AGN. Hence, the systematic omission of such source pairs would bias the test catalog. Since the exact position of the AGN test sources is not important, I did not use the true positions but shifted all AGN randomly by 120 arcsec. As shown in Fig. 4.2, the source density in the vicinity of the shifted sources is uniform as expected.

4.2.2 Construction

The X-ray properties as the X-ray flux, the positional uncertainty, and the existence likelihood are connected through the exposure time and the number of detected source counts. Therefore, I started by randomly selecting an exposure time and an X-ray flux of one of the main RASS sources for each of the 76 million stellar and 1.6 million *Gaia*-AGN sources. Since optically brighter sources also tends to have higher X-ray fluxes, I adopted an X-ray flux of one of the RASS sources with a similar source magnitude as the test source sampling from the stellar RASS sources of Freund et al., submitted (see Sect. 4.1) and the sample of extragalactic RASS sources as described in Sect. 4.2.1. Given the exposure time and the X-ray flux, I derived the number of detected source counts. Since some sources obtained an unrealistic small number of counts, I estimated the detection limit from the exposure time and excluded all test sources constructed below the detection limit. The existence likelihood, the accuracy of the flux measurement, and the position depend on the number of counts, and therefore, I randomly selected these quantities of a RASS source with a similar number of counts. Then, I randomized the flux values and positions of the test sources applying a Gaussian and spherical Gaussian distribution, respectively, to simulate the uncertainties in the measurements of the X-ray sources.

For the creation of the final test catalog, I selected about 20000 and 40000 of the 76 million stellar and 1.6 million extragalactic sources, while excluding those sources, for which the constructed X-ray flux is below the detection limit. However, the detection limit biased the sample to bright and high active sources because many artificial sources fall below the detection threshold by construction.

To conserve the properties of the RASS catalog, I sorted the test sources in count, activity, color and distance bins and randomly selected the correct number of sources bin-wise. However, the test catalog lacks some stellar test sources and AGN with a low activity and a small number of counts because these bins do not contain enough sources. To obtain spurious sources, I assigned new, random positions to a fraction of the selected sources depending on their existence likelihood as given in Table 1 of Boller et al. (2016).

4.2.3 Results

The resulting test catalog contains 15 197 stellar sources, 30 607 extragalactic objects, and 13 539 spurious sources, hence, about 23 % of the test sources are spurious. This is a smaller fraction than expected for the RASS catalog, where 30 % spurious detections are expected (Boller et al., 2016). The discrepancy has two reasons; first, the binned representation of Table 1 of (Boller et al., 2016) leads to a lower estimate of spurious detections than stated in the text of the paper. Furthermore, the RASS sources with a large positional uncertainty that are excluded from the main catalog generally have a smaller number of detected counts and a lower existence likelihood. In Fig. 4.3 I compare different properties of the test catalog with the real RASS sources. Overall, the properties of the test sources agree very well with the RASS catalog. The test sources tend to be located in less dense regions, and furthermore, the stellar test sources are slightly fainter than the real stellar RASS sources.

I applied the same identification procedure to the test catalog as to the RASS catalog (see Sect. 4.1) with the test catalog having the advantage that the correct identification is known for every source. From the test catalog, it is very obvious that the stellar sources cannot reliably identified solely from the geometric matching information. The correct identification is for only 17 % of the stellar test sources the nearest neighbor and for 21 % of the sources, the correct identification is not among the top ten nearest neighbors. Therefore, I created a Bayes map with the same criteria as described in Freund et al., submitted (see Sect. 4.1). Due to the smaller size of the test catalog, its training set contains only 155 sources. When combining the geometric matching properties with the Bayes factor, the best counterpart of 96 % of the stellar test sources is the correct identification. However, this fraction might be lower for the real RASS catalog because the stellar test sources contain less binaries, for which the component that emits the X-ray cannot be reliably identified with the angular resolution of *ROSAT*.

In Fig. 4.4 I compare the true completeness and reliability of the test catalog identifications with the values estimated as described in Equation 15 and 16 of Freund et al., submitted (see Sect. 4.1). With about 93.6 %, the completeness and reliability of the identification for the test sources agrees well with the real RASS sources, although the intersection of completeness and reliability is reached at a higher p_{stellar} cutoff value. The estimated and true values agree overall well; the reliability fits within ~ 0.7 pp and the true completeness is about 2.4 pp higher than the theoretical estimation. The discrepancy might be caused by the small number of training set sources. As for the RASS sources, I also applied the matching procedure to randomly shifted test sources that have only spurious associations by definition. The so obtained reliability agrees almost perfectly with the true values confirming that the reliability can be well estimated by the identification of shifted sources.

4.3 Construction and influence of the Bayes map

Bayes maps are applied in Freund et al., submitted (see Sect. 4.1) and Schneider et al. (2021) (see Sect. 5.1). Due to the Bayes map, counterparts with properties typical for stellar X-ray sources are upweighted and matches with properties that are more often found for random associations are downweighted.

In this section, I will discuss in more detail how the Bayes map is constructed and how much the details of the estimation of the Bayes factors influence the results. For this purpose, I apply Bayes maps as a function of the X-ray to G band flux ratio and the $BP - RP$ colors because these properties do not much vary with Galactic latitude, however, the conclusions also hold for Bayes maps constructed from other properties.

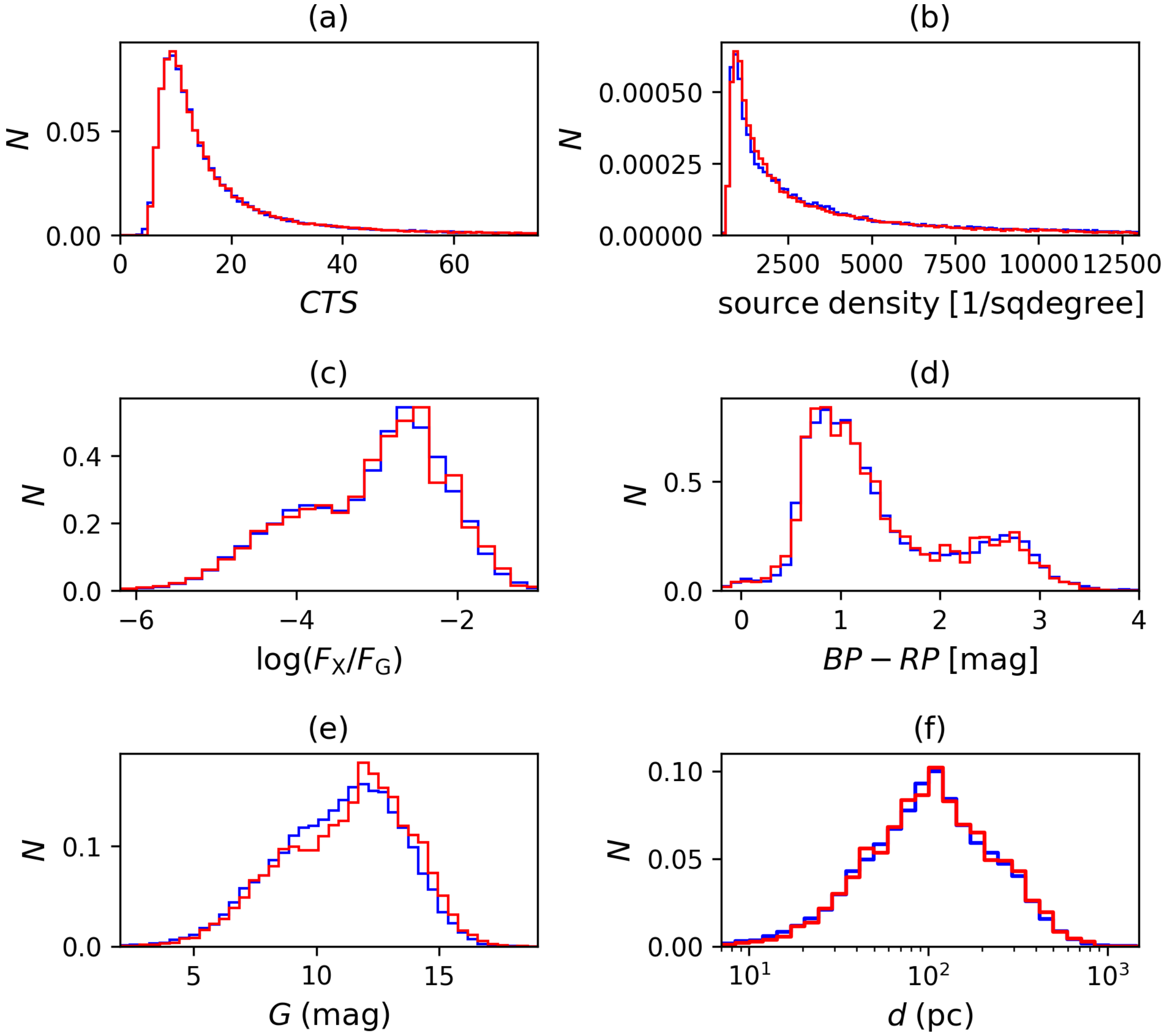


Figure 4.3: Comparison of the properties of the test sources and the real RASS sources. Panel a and b show the number of counts and source density of the whole main RASS catalog (blue) and all test sources (red). In panel c - f, the X-ray to G band flux ratios, BP-RP colors, G band magnitudes, and source distances are compared for the stellar RASS (blue) and stellar test sources (red).

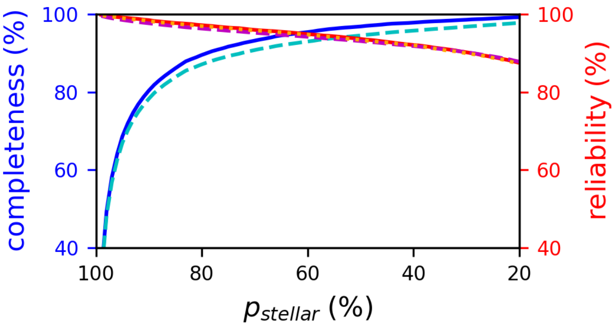


Figure 4.4: Completeness and reliability of the identifications of the test catalog as function of the stellar probability. The blue and red solid curves show the true completeness and reliability, while the dashed cyan and magenta lines indicate the values estimated from the stellar probabilities. The dotted orange line shows the reliability expected from shifted test sources.

4.3.1 Construction

The Bayes map compares the properties of true identifications with spurious associations. To obtain unbiased true stellar counterpart, I selected sources with a high geometric matching probability. Still, a small fraction of the selected sources is spurious or the counterpart is the correct identification but of a source type that I do not seek to identify (e.g. a compact object). Therefore, I excluded sources located below the main sequence or objects with an X-ray luminosity or X-ray to G band flux ratio too high for coronal X-ray emitters (see Freund et al., submitted (Sect. 4.1) and Schneider et al. (2021) (Sect. 5.1) for details). The resulting training sets for the RASS and eFEDS catalog are shown in the middle panels of Fig. 4.5. Both distributions are very similar; F-, G-, and early K-type sources are found from as low as $\log(F_X/F_G) \approx -6$ up to the saturation limit, while nearly all M-type stars are located close to the saturation limit. The eFEDS training set contains more late type stars and sources with a low X-ray to G band flux ratio due to its lower detection limit. In the training set of the RASS catalog also early type objects are found that do not exist in the eFEDS field.

To access the properties of spuriously identified background sources, I performed a cross-match of the optical stellar candidates with sources randomly shifted between 10 and 20 arcmin and by a random angle from the true X-ray position. Due to the shifting, all counterparts are spurious by definition but the large scale spatial distribution of the X-ray sources is conserved. In the top panels of Fig. 4.5 I show the $\log(F_X/F_G)$ vs $BP-RP$ distribution of about 100 000 counterparts to the shifted sources. These sources are generally much fainter than the training set sources, and hence, their X-ray to G band flux ratio is higher. The difference is larger for the RASS sources because the eFEDS sources have on average lower X-ray fluxes due to their lower detection limit, and hence, the F_X/F_G values are lower. Also the shape of the distribution differs because the high Galactic latitude eFEDS field contains different source types than the all-sky RASS catalog.

From the distributions of the training and control sets, I estimated the probability density functions (PDF) f_{train} and f_{spurious} applying a Gaussian kernel density estimator (Rosenblatt, 1956; Parzen, 1962) with a rather large bandwidth compared to Scott's rule (Scott, 1992). The Bayes map is then estimated by dividing f_{train} through f_{spurious} so that the Bayes factor is larger than unity for parameters where the true identifications of the training set have a larger likelihood. In regions where many spurious associations of the control set are found but only a few training set sources, the Bayes factor is < 1. The different sizes of the training and control sets do not influence the results because the PDFs are normed. Since the Gaussian kernel does not reach zero, the Bayes factor would go to zero or infinity in regions not populated by the training or control set. Therefore, I added a small constant to the PDF that does not significantly change the Bayes factor in well populated regions but causes the Bayes factor to go to one in unpopulated areas. The influence of the chosen constant and the bandwidth is discussed in Sect. 4.3.2.

In the bottom panel of Fig. 4.5 I show the resulting Bayes maps for the RASS and eFEDS catalog. Both maps are overall similar with low X-ray to G band flux ratios being upweighted

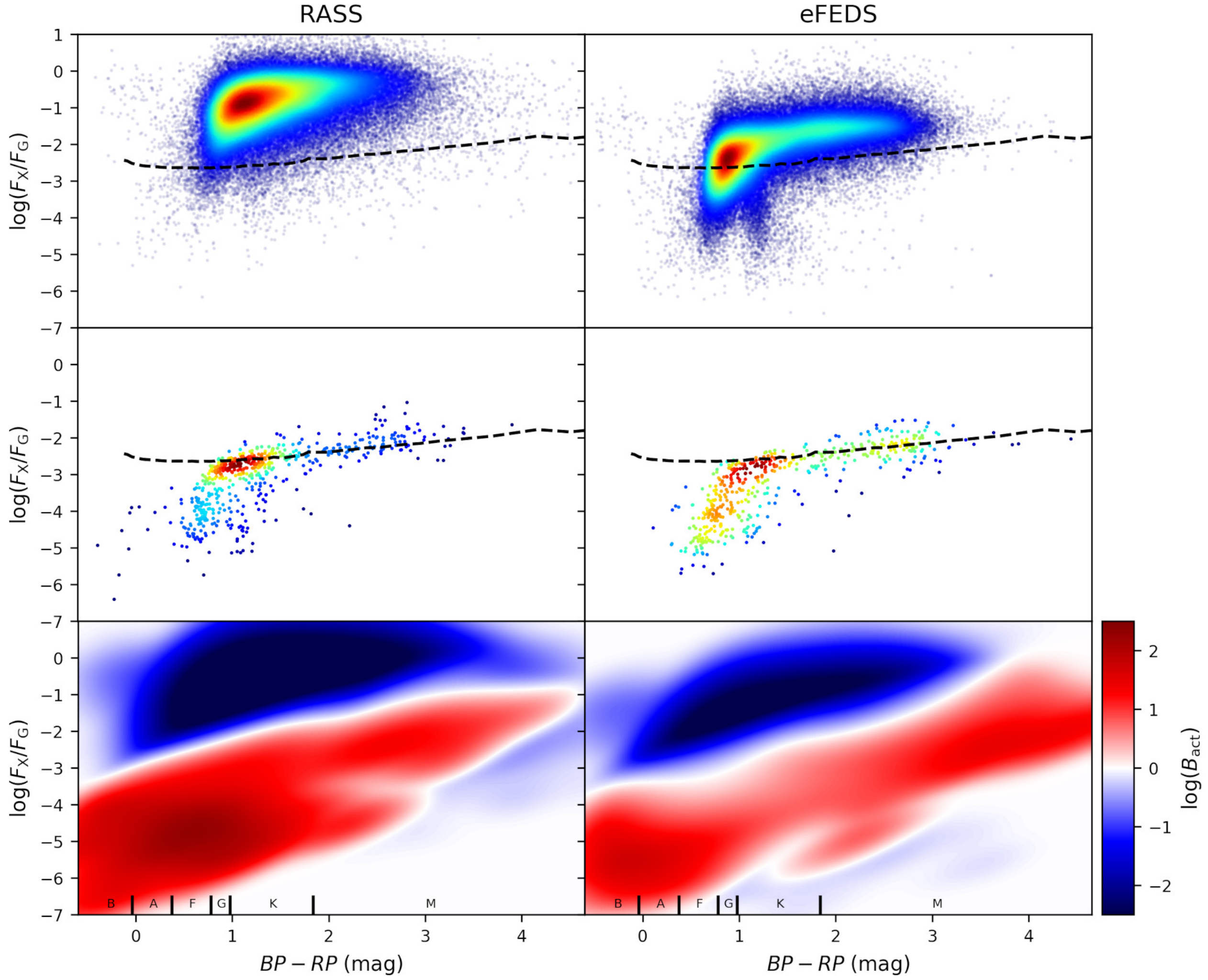


Figure 4.5: Construction of the Bayes map. The panels from top to bottom show the control set, training set, and the resulting Bayes map of the RASS (left) and eFEDS (right) sources. The color of the control and training sets scale with the density of the sources, while for the Bayes maps, the color indicates the strength of the up (red) and down (blue) weighting. The dashed lines indicate the nominal saturation limit at $\log(F_X/F_{\text{bol}}) = -3$.

with Bayes factors > 1 and regions with high F_X/F_G values having Bayes factors < 1. The flux ratio of the dividing line between up and down weighted areas increases for later spectral types. Due to the limited sensitivity of the X-ray surveys, the training and control sets do not detect M-type stars at low flux ratios, and hence, the Bayes factor is near unity in this region. For the eFEDS sources, the overlap between the training and control set is larger than for the RASS sources because of the lower detection limit, and therefore, the Bayes factors for the eFEDS catalog are generally smaller and the weighting weaker. This is the reason why the completeness and reliability of the RASS identifications are higher than for eFEDS, although the eFEDS sources have much better positional accuracies (cf. Freund et al., submitted (Sect. 4.1) and Schneider et al. (2021) (Sect. 5.1)).

4.3.2 Influence

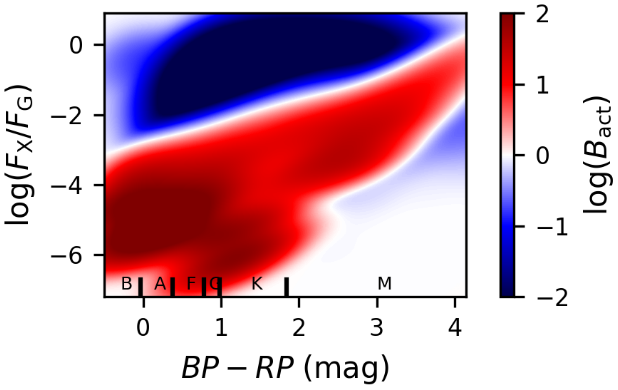
In the last section, I described that I applied a rather large bandwidth to the estimation of the PDF for the training and control set and that I added a small constant. To check how these details of the estimation of the Bayes map influence the identification results, I applied different Bayes maps to the test catalog described in Sect. 4.2.

In Fig. 4.6 I show three different Bayes maps that are constructed from the same training and control sets but with different parameters used to derive the PDF. The first Bayes map is estimated with a large bandwidth and a constant as described in Sect. 4.3.1, I refer to this map as the default Bayes map in the following. Here, the Bayes factor in the lower right part is near unity because such sources are very rare in the training and control set. The strength of the up and downweighting decreases for very low and high X-ray to G band flux ratios, because of the small number of training and control set sources but without physical reasons. For the Bayes map shown in Fig. 4.6(b), I did not add a constant to the PDF. As a consequence, the Bayes factor goes to very high or low values at the edges of the map. For the third map, I adopted a small bandwidth, hence, the Bayes factors are less smoothed and vary stronger sometimes due to individual sources in the training or control set.

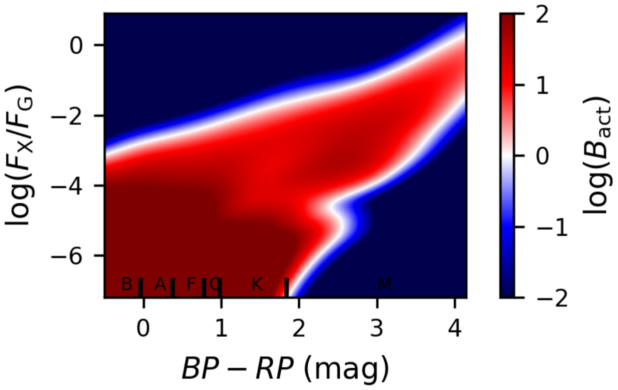
In Fig. 4.7 I compare the identifications obtained with the different Bayes maps. Each Bayes map favors somewhat different sources as most reliable counterparts but the deviation decreases for samples that recover the true number of stellar test sources. The strongest deviations are observed for the spurious associations. The constant added to the PDF has a minor effect on the identifications; the samples of the 5 000 best counterparts differ by 2.4 % and the 15 000 best counterparts agree within 0.4 %. The influence of the bandwidth is stronger, here, the 5 000 and 15 000 best counterparts identified with a large and small bandwidth differ by 12 % and 5 %, respectively. The best results are obtained by the default map.

I also applied the Bayes map constructed for the eFEDS catalog shown in Fig. 4.5 to the test catalog and compare in Fig. 4.8 the so obtained identifications with the results from the default Bayes map of the test catalog. The samples differ with 28 % for the 5 000 best counterparts and 16 % for the 15 000 best counterparts much stronger than for the different estimations of the PDF. Since the weighting of the eFEDS Bayes map is weaker, this maps leads to a higher fraction of spurious associations.

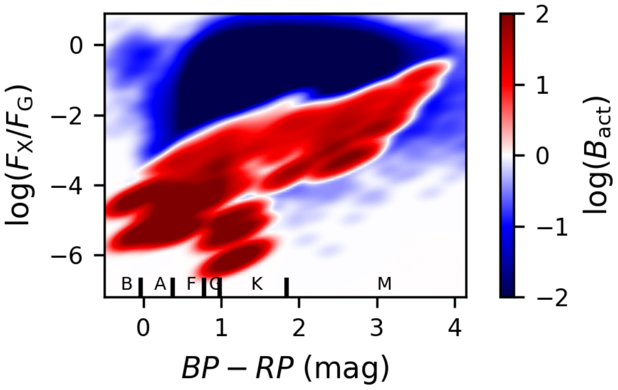
Yet in summary, I conclude that the way of estimating the PDF has a minor influence on the results, although the differences in Fig. 4.6 are apparently large. The regions well populated by the training and control set are not much affected by the details of the estimation. Instead, the Bayes factors strongly differ in regions where only a few training and control sources are found and the Bayes map is poorly determined. However, the number of matches with such properties is also small. On the other hand, the results of the identification substantially change when a Bayes map of sources with other properties is adopted, e.g., the Bayes map of the eFEDS catalog that has a lower detection limit. Here, the weighting is changed in regions where many counterparts are located, and hence, the influence on the identifications is large. As a con-



(a) default



(b) no constant



(c) small bandwidth

Figure 4.6: Comparison of different Bayes maps for the test catalog. The Bayes map in the top panel is constructed with my default parameters, in the middle panel a Bayes map without constant is shown, and for the bottom panel a small bandwidth is applied.

sequence, different Bayes maps are needed for every X-ray survey.

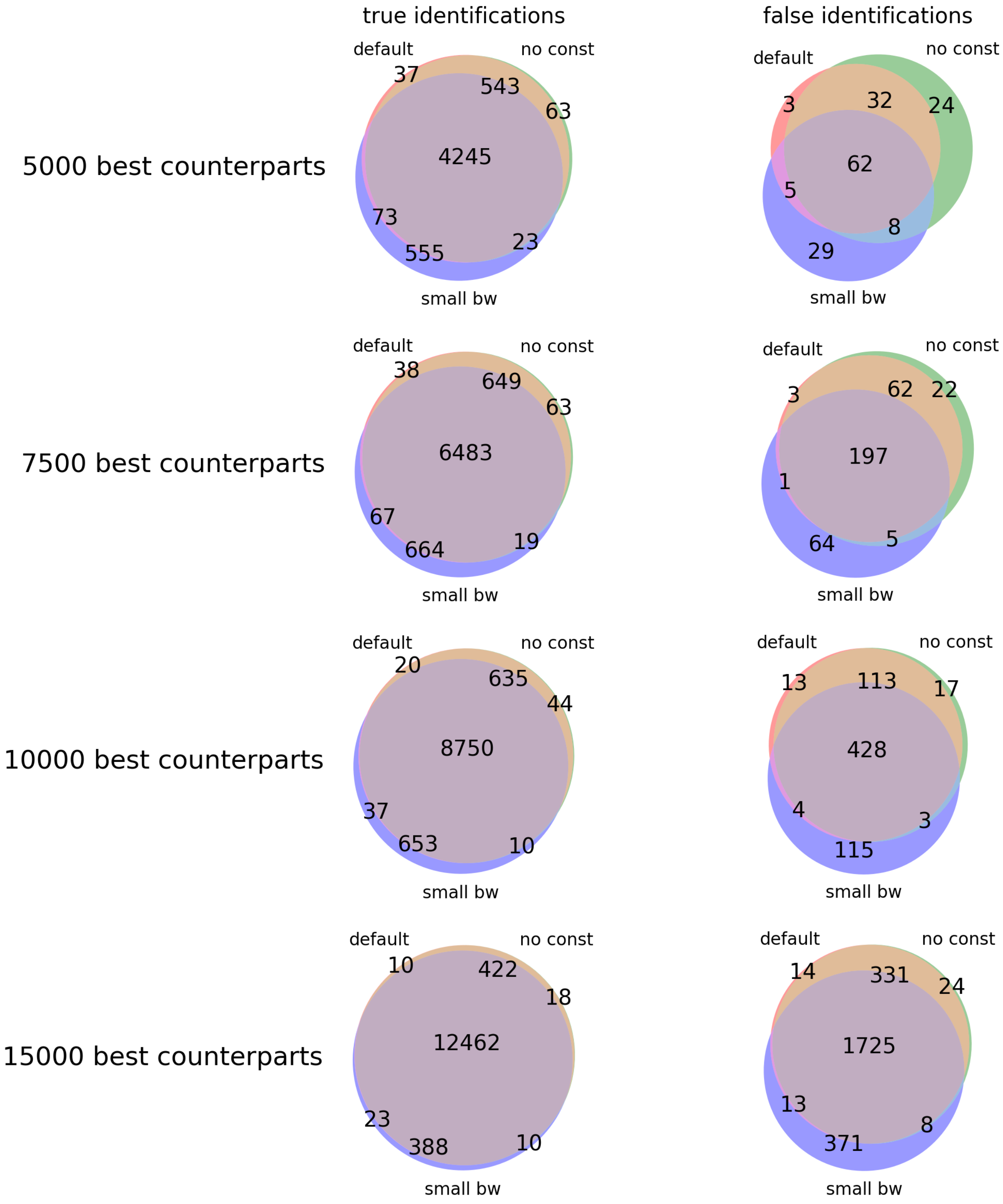


Figure 4.7: Comparison of the 5 000, 7 500, 10 000, and 15 000 best counterparts identified with the Bayes map shown in Fig. 4.6. In the left and right row the correct and false identifications are shown. The gray area and the corresponding number indicate the sources consistently identified by all Bayes maps. The red, green, and blue areas show the number of sources only identified by the default Bayes map, with no constant, and a small bandwidth, respectively. The sources not identified by the default map, with no constant, and a small bandwidth but by the other two Bayes maps are shown by the light blue, violet and orange regions, respectively.

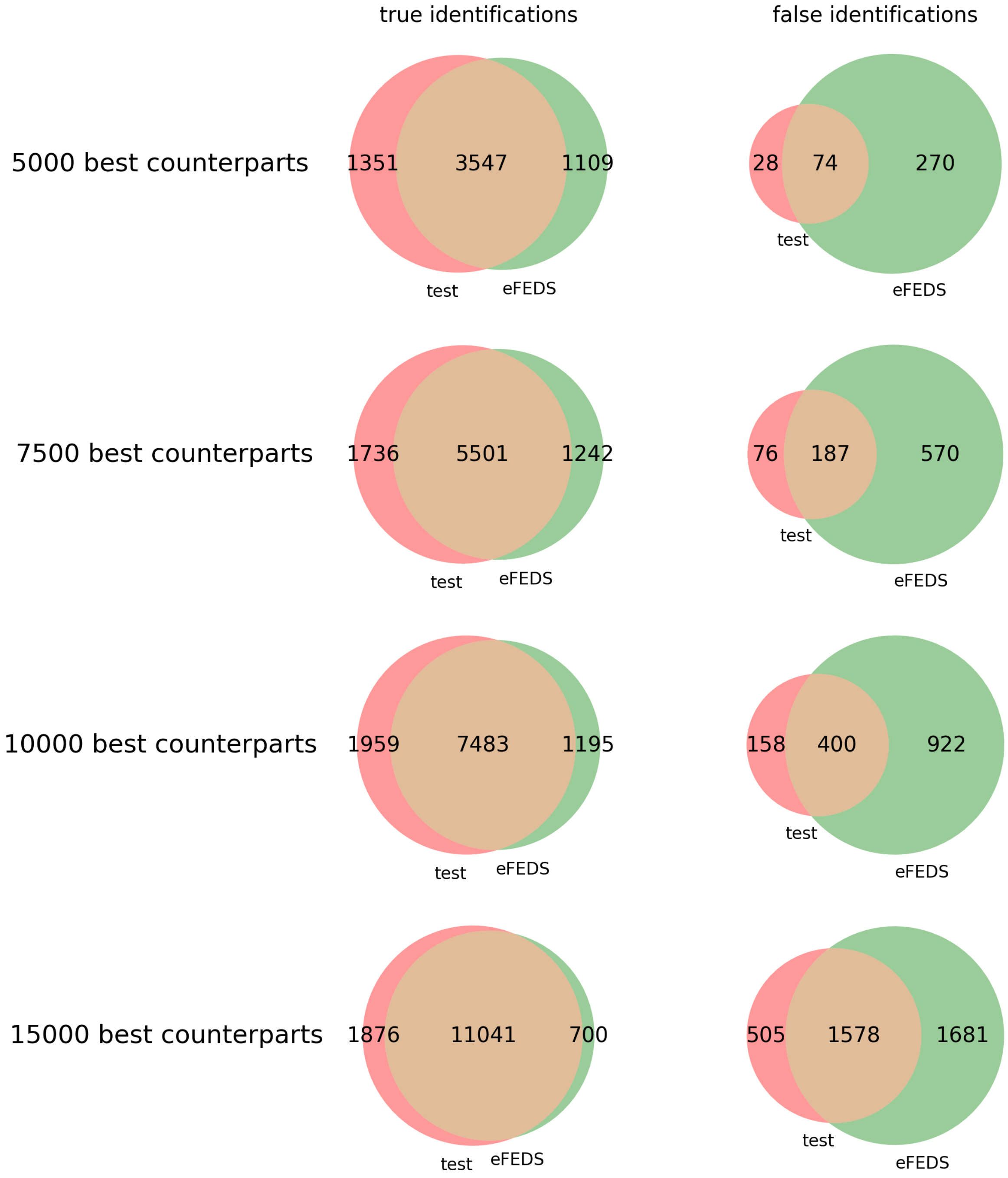


Figure 4.8: Comparison of the 5 000, 7 500, 10 000, and 15 000 best counterparts identified with the default Bayes maps constructed from the test and eFEDS sources. In the left and right row the correct and false identifications are shown. The yellow area and the corresponding number indicate the sources consistently identified by both Bayes maps, while the counterpart only identified by the Bayes map of the test and eFEDS catalog are shown in the red and green area, respectively

Chapter 5

Comparison to other identification algorithms

The results of the Bayesian matching procedure can be compared to other identification approaches. In Schneider et al. (2021) (Sect. 5.1) the stellar identifications of the eFEDS sources obtained with the Bayesian and a Support Vector Machine (SVM) approach are compared. Both methods provide a similar completeness and reliability of almost 90 % and the two samples of stellar identifications are in agreement, incidentally, also of about 90 %. Although the eFEDS sources have better positional accuracies (typically 3 arcsec) than the RASS sources, the quality of the identifications is slightly lower than for the RASS because the eFEDS sources are fainter and the properties of the true identifications stronger overlap with the background sources (cf. Sect. 4.3). In advantage to the SVM approach, the Bayesian framework directly provides probabilities for each counterpart, however, for the SVM approach, the probabilities can be estimated from the expected number of stellar sources and counterparts to randomized sources.

The article of Schneider et al. (2021) is accepted for publication in Astronomy & Astrophysics. P. C. Schneider and I are responsible for the SVM and Bayesian approach, respectively. The source selection and the interpretation was reached in collaboration with about equal contributions. The other coauthors provided contributions that cannot be named individually.

Another identification approach for the eFEDS sources is provided by Salvato et al. (2021) in Sect. 5.2. In contrast to the methods presented in Schneider et al. (2021) (Sect. 5.1), the procedure of Salvato et al. (2021) is not specialized for stars but searches for counterparts to all eFEDS sources independently of their source type but with a special emphasis on AGN. The agreement between the Bayesian method of Schneider et al. (2021) and Salvato et al. (2021) is almost 90 %.

The article of Salvato et al. (2021) is submitted for publication in Astronomy & Astrophysics. I provided contributions to the comparison with the Bayesian method of Schneider et al. (2021) as described in Sect. 4.5. of Salvato et al. (2021) (Sect. 5.2).

5.1 Publication: The eROSITA Final Equatorial-Depth Survey (eFEDS): The stellar counterparts of eROSITA sources identified by machine learning and Bayesian algorithms

Astronomy & Astrophysics manuscript no. main
June 29, 2021

The eROSITA Final Equatorial-Depth Survey (eFEDS)

The stellar counterparts of eROSITA sources identified by machine learning and Bayesian algorithms

P. C. Schneider[1], S. Freund[1], S. Czesla[1], J. Robrade[1], M. Salvato[2], and J. H. M. M. Schmitt[1]

[1] Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany e-mail: `astro@pcschneider.eu`
[2] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, 85748 Garching, Germany

Received ?; accepted ?

ABSTRACT

Stars are ubiquitous X-ray emitters and will be a substantial fraction of the X-ray sources detected in the on-going all-sky survey performed by the eROSITA instrument aboard the Spectrum Roentgen Gamma (SRG) observatory. We use the X-ray sources in the eROSITA Final Equatorial-Depth Survey (eFEDS) field observed during the SRG performance verification phase to investigate different strategies to identify the stars among other source categories. We focus here on Support Vector Machine (SVM) and Bayesian approaches, and our approaches are based on a cross-match with the Gaia catalog, which will eventually contain counterparts to virtually all stellar eROSITA sources. We estimate that 2060 stars are among the eFEDS sources based on the geometric match distance distribution, and we identify the 2060 most likely stellar sources with the SVM and Bayesian methods, the latter being named HamStars in the eROSITA context. Both methods reach completeness and reliability percentages of almost 90 %, and the agreement between both methods is, incidentally, also about 90 %. Knowing the true number of stellar sources allowed us to derive association probabilities p_{ij} for the SVM method similar to the Bayesian method so that one can construct samples with defined completeness and reliability properties using appropriate cuts in p_{ij}. The thus identified stellar sources show the typical characteristics known for magnetically active stars, specifically, they are generally compatible with the saturation level, show a large spread in activity for stars of spectral F to G, and have comparatively high fractional X-ray luminosities for later spectral types.

Key words. stars: activity, stars: X-rays, stars: coronae

1. Introduction

Stars with convective envelopes, that is, with stellar masses between 0.08 and 1.85 $M_\odot$ (spectral types M to mid A) show magnetic activity and possess a corona emitting soft X-rays ($\lambda \sim 1-100$ Å). Stars of the same mass, however, can have very different X-ray properties, primarily depending on the stellar rotation period. The largest ratios between X-ray (L_X) and bolometric luminosities (L_{bol}) are observed for rapidly rotating young stars, which show L_X/L_{bol} ratios close to the so-called saturation limit of $\log L_X/L_{bol} \sim -3$ (Vilhu 1984; Pizzolato et al. 2003). On the other hand, old slowly rotating stars may have $\log L_X/L_{bol} \sim -8$ (see reviews by Güdel 2004; Testa et al. 2015). Since stars spin down with age, the stellar X-ray luminosity also declines with age (Skumanich 1972). Therefore, X-ray surveys are most sensitive to young stars and contain comparatively few old stars, hence, the parameter space of stellar activity is very unevenly sampled.

In addition to this inherent bias toward active stars, existing stellar samples with well characterized X-ray properties are relatively small, typically $\lesssim 1\,000$ objects (Schmitt & Liefke 2004; Wright et al. 2011; Freund et al. 2018), compared to other stellar samples with hundreds of thousands of stars such as RAVE (Steinmetz et al. 2006), the Gaia-ESO survey (Gilmore et al. 2012), or even billions of stars (Gaia Collaboration et al. 2016). The eROSITA all-sky survey, described in Predehl et al. (2021) and designated as eRASS:8, is expected to provide detections of almost 10^6 stars (see Merloni et al. 2012), thus bringing the sample count of X-ray emitting stars on par with other samples.

To harvest the full potential for stellar science, one needs to identify the stars among the other X-ray emitting objects in the eROSITA source list, that is, a classification task. The final data of the eROSITA all-sky survey will not be available before 2024, however, the eROSITA Final Equatorial Depth Survey (eFEDS) already provides the X-ray sensitivity expected after the completion of the all-sky survey for a field of $\sim$ 140 square degrees (Brunner et al. 2021; Salvato et al. 2021, submitted to A&A). Hence, eFEDS provides an excellent opportunity to develop the methods required for the identification task.

The classification of large numbers of objects into different categories based on their measured properties is an old task and has been approached by different mathematical methods such as frequentist (e.g., Fischer 1938) or Bayesian procedures (Binder 1978). Nowadays, machine learning (ML) approaches are also popular thanks to improving algorithms, increasing computing power, and large datasets. These techniques are now regularly applied in the astrophysical context (e.g., Marton et al. (2019); Vioque et al. (2020), and Melton (2020) used ML techniques to identify young stars). Here, we present a Support Vector Machine (SVM) and a Bayesian method to identify stellar X-ray sources within the eROSITA source catalog.

The paper is structured as follows. We translate the task of identifying the stars into a classification problem in sect. 2, present the SVM and Bayesian approaches in sects. 3 and 4,

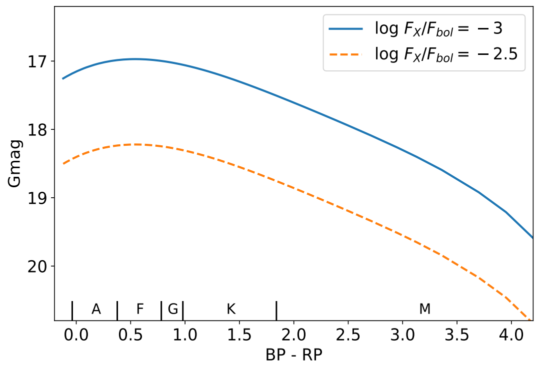


Fig. 1. Limiting G magnitude for stars emitting at X-ray activity levels of $\log L_X/L_{bol} = -2.5$ and -3 assuming the average eFEDS exposure depth.

compare their results in sect. 5, and provide our conclusions as well as the outlook in sect. 6.

2. The identification task

To identify and characterize the stellar eROSITA sources, we need information from other wavelengths in addition to the X-ray data. Specifically, the identification of the stellar content in eROSITA is based on matching the eROSITA sources to a catalog containing only eligible stellar counterparts, i.e., stars that may emit X-rays at flux levels detectable by eROSITA. An eROSITA source i is classified as stellar if an association between eROSITA source i and a counterpart j exists that has properties compatible with the hypothesis that counterpart j is responsible for the X-ray emission, that is, if one or more reasonable stellar counterparts $(j_1, \ldots, j_n)$ exist so that the association $i \leftrightarrow j_1$ (or $i \leftrightarrow j_2$, etc.) is likely while taking the possibility of chance alignments into account, too.

This approach differs from many other catalog cross-matching approaches in the sense that a complete identification of the eROSITA detections is not required: stars and non-stars are the only two relevant source categories for us, and only for objects in the first category do we attempt to find the correct counterpart (the problem of the identification, characterization and classification of the full eFEDS sample is discussed in Salvato et al., subm.). The non-stars category includes other source categories, mainly AGN (Liu et al., submitted) and nearby galaxies (Vulic et al., submitted), but also spurious X-ray detections, for example, due to background fluctuations. Objects in the non-stellar category are here treated as random associations and considered in a statistical sense. In summary, the categorization of the eROSITA sources is deferred to classifying the associations $i \leftrightarrow j$ between eROSITA sources (i) and eligible stellar counterparts (j) as probable with respect to the alternative that the association is spurious, i.e., that object j in the match catalog is not responsible for X-ray source i.

In this scheme, the match catalog containing the eligible stellar counterparts is crucially important for the classification and must include counterparts to virtually all stellar eROSITA sources but may lack any other source category. The saturation limit for stars of $\log L_X/L_{bol} \approx -3$ (Pizzolato et al. 2003; Wright et al. 2011) implies a limiting optical magnitude for reasonable stellar counterparts at any given X-ray sensitivity (in larger samples, the nominal L_X/L_{bol}-values of some individual stars may exceed $\log L_X/L_{bol} \approx -3$, but excursions toward significantly higher values are very rare). Figure 1 shows the *Gaia* magnitude that a star emitting X-rays at the saturation limit will have for the sensitivity of eFEDS, which is also approximately the limit for the eROSITA survey after eight passes (eRASS:8, see Predehl et al. 2021). In particular, virtually all stellar counterparts of eROSITA sources (eFEDS and eRASS) are expected to have G-magnitudes brighter than the detection limit of *Gaia* and, thus, will be eventually included in the *Gaia* catalog (Gaia Collaboration et al. 2016). The *Gaia* catalog has the additional benefit of being all-sky and future data releases will improve the completeness to levels sufficient for identifying stars in the full eRASS survey; the current data release (EDR3) is already complete for stars between G=12 mag and 17 mag (Gaia Collaboration et al. 2020). Given these beneficial properties, we chose to use the Gaia catalog as our match catalog (and magnitudes are therefore in the VEGAMAG system).

Our identification scheme aims to identify coronal emitters in quiescence. Stars occasionally flare, which can elevate the L_X/L_{bol} quite significantly (see Boller et al. 2021, subm., for examples). Currently, stars undergoing flares during the eROSITA observations may show, depending on the specific star and flare, L_X/L_{bol}-values that are so high ($L_X/L_{bol} \gg -3$) that they are deemed incompatible with stars in our schemes and, thus, may be misclassified as non-stellar (associations with L_X/L_{bol}-values somewhat above -3 are typically still classified as stellar, see Fig. 6). The methods discussed here do not attempt to correct for the effects of flaring, because the observing sequence of the eROSITA all-sky survey differs from that of eFEDS. For eRASS, flares will be easily detected in the survey as each object is scanned multiple times. Also, we concentrate on coronal X-ray sources and other galactic X-ray sources exist, which include genuine stars (e.g., CVs). The origin of the X-ray emission in these objects, however, differs from the coronal emission seen in "normal" stars and they typically have high fractional X-ray luminosities. Therefore, they are also unlikely to be classified as stellar with our methods.

2.1. Input catalogs and data screening

The eROSITA and *Gaia* catalogs contain entries that are very unlikely to describe stars detected by eROSITA and we applied a number of filter criteria before performing the stellar identification to remove such catalog entries.

2.1.1. Stellar candidates in the eROSITA eFEDS catalog

We used eROSITA detections from the main eFEDS catalog, which contains sources detected in the energy range between 0.2 and 2.3 keV (see Brunner et al. 2021) and applied the following filter.

1. No significant spatial extension (`EXT_LIKE`< 6),
   Extended sources are unlikely to represent stars unless they are a blend between two (or more) sources (expected to be very rare).
2. positive `RADEC_ERR`.
   There is a small number of eROSITA sources for which the source detection algorithm failed to calculate reasonable positional uncertainties. Because the positional uncertainty is a key value for our matching algorithm, we ignored sources with negative or zero `RADEC_ERR` entries.

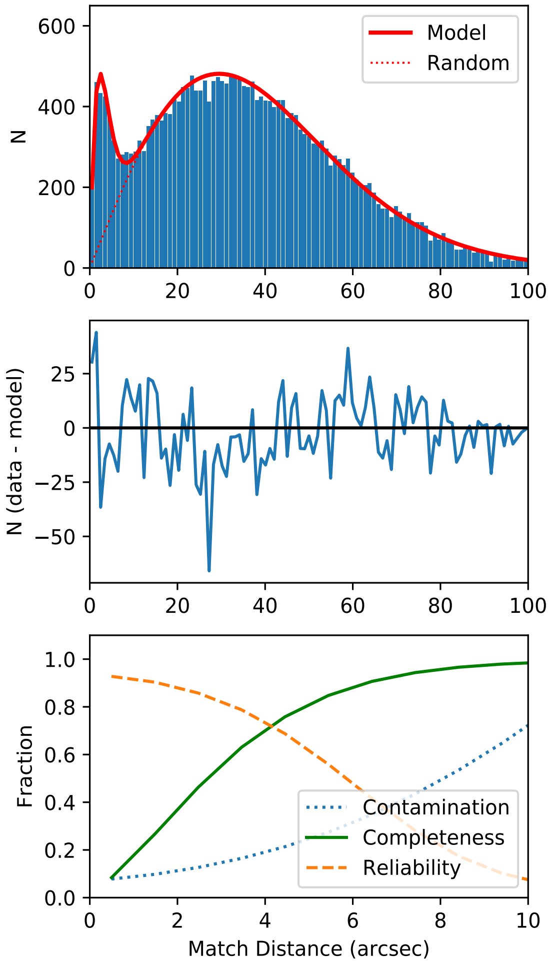


Fig. 2. Top: Nearest neighbor distance distribution between eROSITA and eligible Gaia sources. Also shown are the best-fit model using Eq. 4 and the expected number of random matches at each distance. **Middle** Difference between model and data. **Bottom**: Completeness, reliability, and contamination for samples including stars up the given match distance.

Applying these criteria resulted in $N_X = 27\,369$ eROSITA detections, which constitute the X-ray input catalog abbreviated with $\mathcal{X}$ in the following.

2.1.2. Eligible *Gaia* counterparts

For the *Gaia* sources, we propagated coordinates from the *Gaia* EDR3 reference epoch (2016.0) to the epoch of the eFEDS observation (2019.85). Among the *Gaia* catalog entries, we selected those sources fulfilling the following conditions to be considered eligible counterparts to eROSITA detected stars.

1. Gaia sources above a certain brightness limit given by the X-ray sensitivity,

   The eROSITA X-ray sensitivity limit combined with the range of L_X/L_{bol} ratios of stars (see sect. 1) implies that eligible counter parts to eROSITA detections must have $G < 19$ mag (for spectral types A to M, see Fig. 1).
2. *Gaia* sources must have a three sigma significant parallax measurement (`parallax_over_error`> 3),

   Gaia entries that do not have a significant parallax measurement are mostly extra galactic or galactic objects so distant ($d \gg 1$ kpc) that they unlikely represent eligible counter parts to eROSITA detected stars.
3. *Gaia* sources that fall into a rather generously defined region of the *Gaia* color-magnitude diagram (see App. A).

   This region includes young (1 Myr) and old (10 Gyr) stars with some margin to account for measurement errors. Regions in color-magnitude space where other objects such as cataclysmic variables (CVs) may be found are excluded (below the main sequence). We consider the stellar properties for objects outside this region to be too poorly known for our further analysis.

We did not filter the *Gaia* sources for RUWE, because we found that the sources with larger RUWE values[1] have perfectly acceptable properties and suspect that the selection bias towards active stars and binaries or multiples may, at least partially, influence the astrometric solution leading to less than perfect RUWE values. In total, $N_G \approx 4 \times 10^5$ Gaia EDR3 counterparts fulfill our quality criteria (called $\mathcal{G}$ in the following).

2.2. *The catalog fraction*

Among the properties describing any given eROSITA to Gaia association $i \leftrightarrow j$, the matching distance is special: Measuring the sky positions with very high precision unambiguously informs us about the correct counterpart, and if such a counterpart does exist in the match catalog, because sources with too large matching distances have negligible likelihoods to be the correct counterpart independent of all other parameters provided that source confusion is negligible.

This peculiarity of the sky positions provides a particularly beneficial information for the matching procedure, namely an estimate for the number of real matches, i.e., the fraction of eROSITA sources with a counterpart in the match catalog (the screened Gaia catalog), which we call the geometric catalog fraction (CF) in the following. The CF can be derived from the measured on-sky distances between the eROSITA sources and their nearest Gaia match, because the probability distributions to measure a particular match distance is known a priori for real and random associations. The distribution in match distance depends only on the positional uncertainty of the eROSITA source σ_i for real associations (in our context, Gaia sources have negligible positional uncertainties) and on the local sky density of eligible sources in the match catalog for random associations, respectively. Both properties can be measured independently of the nature of the eROSITA sources and its membership in the match catalog.

The CF is a sample property affecting all associations equally; it cannot be used to select any particular association over another association. For example, a CF of 0 % would imply that even near perfect positional matches cannot be considered real while for a CF of 100 %, large match distances are perfectly acceptable, because the correct counterpart must be among the

---

[1] A RUWE value of 1.4 was suggested for single stars within 100 pc, see `http://www.rssd.esa.int/doc_fetch.php?id=3757412`

candidates and one “just” needs to identify the correct counterpart among the ensemble of candidates.

Figure 2 (top) shows the measured nearest neighbor match distance distribution between eROSITA and Gaia sources for eFEDS. Assuming that the positional errors σ_i and sky densities η_j do not differ between real and random associations, i.e., sample from the same parent population, the match distance distributions are known for real and random associations and the model for the match distance distribution has only one free parameter, namely the CF. In particular, the match distance distribution is

$$d_{real}(r_{ij}, \sigma_i) = \frac{r_{ij}}{\sigma_i^2} e^{\frac{-r_{ij}^2}{2\sigma_i^2}} \qquad (1)$$

for real matches with the match distance r_{ij} for the association between the i-th eROSITA and the j-th Gaia source and the Gaussian positional uncertainty σ_i, which we calculated from `RADEC_ERR` as

$$\sigma_i = \sqrt{\frac{\text{RADEC\_ERR}_i^2 + 0.7^2}{2}} \qquad (2)$$

including a systematic uncertainty of 0.7 arcsec (Brunner et al. 2021). The likelihood to find a random association $i \leftrightarrow j$ within r_{ij} and $r_{ij} + dr$ is proportional to r_{ij} and the local sky density. Specifically, the distribution of match distances towards the nearest random neighbor is described by

$$d_{random}(r_{ij}, \eta_j) = 2\pi r_{ij}\eta_j e^{-\pi\eta_j r_{ij}^2} \qquad (3)$$

where η_j is the local sky density, which we measured from the local neighborhood of any eligible Gaia source. The model for the measured match distance distribution is then

$$d_{measured}(r) = \sum_{ij} \Big(CF\, d_{real}(r, \sigma_i) + (1 - CF)\, d_{random}(r, \eta_j) \Big) \qquad (4)$$

where the summation is over all associations between eROSITA and their nearest Gaia match, that is, the CF describes the relative normalization between real and random associations.

Figure 2 (top and middle) shows that such a model for the match distance distrubution accurately describes the measurements and we found a CF of 7.5 % for eFEDS corresponding to 2060 ± 17 stellar sources. As expected, stars represent only a small fraction of the eROSITA detections in eFEDS. In addition, a thorough statistical treatment in the form of a Bayesian mixture model confirms the above number (to be presented in Freund et al., in prep.). The bottom panel of Fig. 2 shows that almost all real associations have match distances of 10 arcsec or less, which is expected based on a median positional error σ of about 3 arcsec (as the median corrected `RADEC_ERR` is about 4.6 arcsec).

Knowing the CF implies that we know the number of stars (2060), but not which specific eROSITA sources are stellar. Knowing the CF also implies that when a set of $N = 2060$ eROSITA sources is classified as stellar, the number of stars misclassified as spurious N_{missed} (type II error or false negative) and the number of sources erroneously classified as stars $N_{spurious}$ (type I error or false positive) are equal, i.e., completeness and reliability are equal. Using the following definitions

$$\text{completeness} = \frac{N - N_{\text{spurious}}}{N - N_{\text{spurious}} + N_{\text{missed}}} \qquad (5)$$

$$\text{reliability} = \frac{N - N_{\text{spurious}}}{N}, \qquad (6)$$

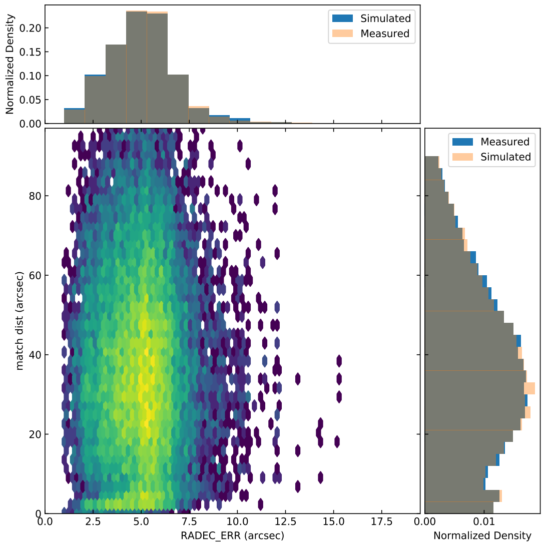


Fig. 3. Measured and simulated match distances for eFEDS. In the histograms, the gray bars indicate the overlap between the measured and simulated samples.

Fig. 2 (bottom) shows that the completeness and reliability are about 70 % accepting associations up to a fixed match distance of $d_{max} = 4$ arcsec, i.e., one would classify the correct number of eROSITA sources as stellar, but almost every third classification would be erroneous. The task at hand is therefore to improve the completeness and reliability of a sample containing the 2060 most likely stellar sources using more information than just the match distance.

3. Method I: Support Vector Machine

To identify the stellar X-ray sources, we now want to classify all associations $i \leftrightarrow j$ between eROSITA X-ray sources and eligible *Gaia* counterparts into the previously defined categories of stellar and non-stellar associations. This binary classification task resembles problems regularly approached with ML techniques.

The feature vector, which contains the properties describing an association $i \leftrightarrow j$ such as the match distance r_{ij}, plays a crucial role for all ML classification methods, raising the question which features should be included in it. In principle, any combination of properties available from the Gaia and eROSITA catalogs may be used to populate the feature vector describing an association. However, the eFEDS dataset is somewhat limited in source numbers for ML methods to extract the relevant properties in the feature vector in addition to fitting or training. Therefore, we opted for an astrophysically motivated choice of properties, which are listed in Table 1. Specifically, we constructed feature vectors with $M_X = 2$ X-ray (F_X, σ) and $M_G = 4$ Gaia features (F_G, plx, η, and $BP{-}RP$) plus the match distance r_{ij}, so that the feature vector is seven dimensional ($\mathbf{p}_{ij} \in \mathbb{R}^{M_X+M_G+1} = \mathbb{R}^7$).

Our choice of features has the highly advantageous property that the probability for a correct classification is a monotonic function in many features listed in Table 1. For example, an association $i \leftrightarrow j_1$ with a Gaia source j_1 being less distant (in pc) is generally the more likely counterpart than a Gaia source j_2 at a larger distance when all other features are equal, because the number of chance alignments increases with distance. Similarly, the higher the optical flux, the higher is the likelihood that the association under consideration is correct, because optically bright sources are rare and, clearly, the smaller the on-sky match distance, the higher is the likelihood of a correct association. In fact, we used the latter property to strongly reduce the number of to-be-classified associations by considering only plausible ones with match distances of up to 60 arcsec.

The monotonic behavior of the feature vector with respect to the association probability makes our classification task ideally suited for a support vector machine (SVM) approach (Cortes & Vapnik 1995). A SVM classifies a feature vector $\mathbf{p}_{ij}$, which in our case characterizes an association, into the two categories of stellar and non-stellar associations. The basic idea of a SVM classifier (SVC) is to use a training sample with labeled data to construct a so-called maximum-margin hyperplane separating the two categories in feature space. The confidence in the classification increases with distance from this hyperplane; samples near the hyperplane have less secure classifications compared to samples far away (in feature space) from the hyperplane. This behavior is well matched to the content of our feature vector and we used the SVC implementation of scikit-learn (Pedregosa et al. 2011) with a polynomial kernel to categorize associations $i \leftrightarrow j$ in the following.

3.1. Training and validation Samples

The quality of the training sample is crucial for the final classification. Among the available information on the associations, the match distance stands out for indicating likely matches independent of any other property. Therefore, we constructed our final training sample in two steps, first, focusing on geometry and, second, incorporating physical properties.

3.1.1. Selection of the training sample with a geometric SVC

In a first step, we constructed a geometric training sample based on the best positional associations and some quality criterium as described in the following. The reliability of an association depends on the positional uncertainty of the X-ray source σ and on the local density of eligible counterparts η, because the expected number of random matches scales as $\eta\sigma^2$, which must be small for a reliable association. Hence, the same match distance may be perfectly acceptable for low sky densities while unacceptable for high sky densities.

To identify the N best geometric associations for the final training sample, we used a geometric SVC with a feature vector consisting of match distance, positional uncertainty, and local sky density. As the geometric distributions for real and random source are known (cf. sect. 2.2) as well as the CF, sampling from these distributions provided us with large ($N \gtrsim 10^4$) geometric training sets. These training data reproduce the geometric properties of the eFEDS field, and here we also have the labels to train the SVC. As an example, Fig. 3 shows the resulting match distance distribution for the geometric training sample, demonstrating the excellent overlap with the observed distribution in the data. The geometric SVC was then trained to classify associations based on geometry allowing a contamination of 5 % spurious associations, which we found to balance sample size and contamination.

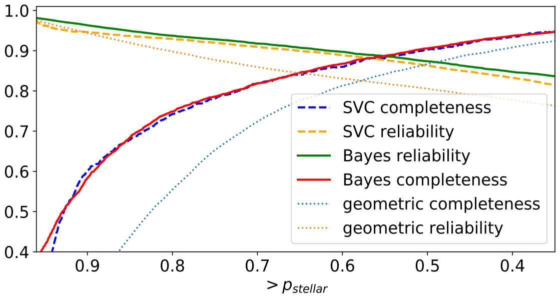


Fig. 4. Completeness and reliability of the stellar identifications for the eFEDS sources as a function of the probability cutoff for the SVM and Bayesian algorithms as well as the purely geometric version.

This geometric SVC was then used on the real data classifying 627 associations as real. As a cross-check, we applied the geometric SVC to randomly shifted eFEDS X-ray sources, which produces only random matches by construction. In this case, the geometric SVC classified 36 associations as stellar, which corresponds to an empirical contamination level of just under 6 % well in line with the expected value.

In a second step, we applied a physical screening of the geometric training sample. In particular, we excluded associations with counterparts yielding $L_X > 2 \cdot 10^{31}$ erg s^{-1} and empirically defined a limit of $\log(F_X/F_G) < (BP - RP) \times 0.655 - 3.22$ for the fractional X-ray fluxes of the training sample objects, larger F_X/F_G-values are well above the saturation limit and unlikely to be produced by stellar coronal X-ray emission in quiescence; only highly active A-type stars may lie above the limit, but are not present in our data anyway. We further excluded a few associations with counterparts brighter than $G = 5.5$ mag, because they potentially cause optical loading (although the association itself may still be the correct one, just the properties could be skewed). Finally, we identified 537 bona fide stellar associations to represent the physical training sample.

This physical training sample contains, by construction, only associations with relatively small match distances, which do not sample the ensemble of expected match distances of true associations well. Therefore, we constructed the final training sample by reasigning new match distances sampling the expected distrubution for real associations, just like for the geometric training sample. This time, however, we kept the X-ray and optical properties of the previously identified associations in the respective feature vectors. The final training sample also includes random associations, which were selected in proportion to the ratio implied by the CF, i.e, the training set reproduces the true ratio between real and spurious associations. The properties of random associations can easily be explored by shifting the eROSITA X-ray sources w.r.t. the background of eligible Gaia counterparts, which is what we did to incorporate them into the final training sample.

3.1.2. Validation sample

The validation sample was constructed similarly as the random associations in the training set (see above), that is, by matching randomly shifted eFEDS sources. This validation sample contains *no* correct associations and all associations classified as stellar by the SVM must be spurious (false positives) by construction. Usually, this would give only part of the desired infor-

Table 1. Properties used for classification

	Name	Abbreviation in text	Description	Unit
1	Angular separation	r_{ij}	Angular distance between eROSITA and Gaia source	arcsec
2	Local sky density	η	Local sky density of eligible Gaia sources	arcmin^{-2}
3	Positional uncertainty	σ	Uncertainty in position for eROSITA source	arcsec
4	Optical flux	F_G	Optical flux in Gaia band	erg s^{-1} cm^{-2}
5	X-ray flux	F_X	X-ray flux for eROSITA source	erg s^{-1} cm^{-2}
6	Gaia color	BP-RP	Color of Gaia source	mag
7	Distance	d	Distance of Gaia source (from parallax)	pc

mation, i.e., would not allow an assessment of the completeness. Having an accurate estimate for the number of stellar eFEDs sources from the CF, however, provides this missing piece of information and the completeness is simply

$$\text{Completeness} = \frac{N - N_{spurious}}{N_{stars}}, \tag{7}$$

where N is the number of eFEDS sources with at least one association classified as stellar and $N_{stars} = 2060$ is the known number of stars from sect. 2.2. Therefore, we do not need a validation sample containing real stellar X-ray emitters to evaluate the completeness and reliability of the classifier.

3.2. Preprocessing

Scales and ranges differ between the features used for the classification, e.g., X-ray fluxes are roughly in the range of 10^{-14} erg s^{-1} cm^{-2} to 10^{-12} erg s^{-1} cm^{-2} while we considered match distances between zero and 60 arcsec. In such cases, it is often recommended to scale each feature to some "standardized" distribution, for example, to normalize to zero mean and standard deviation one. Empirically, however, we found that this does not provide good results for the problem at hand in terms of sample reliability and reliability. Therefore, we opted for individually scaling the features such that numerical values are roughly in the range between 0 and 10 and, e.g., used logarithmic fluxes. The exact scaling values and zero points impact the importance of individual features in the mixed terms of the polynomial kernel and may help, to some degree, to adjust their respective weights in the SVM.

3.3. Optimization goal

Classification tasks often imply a tradeoff between completeness and reliability, e.g., one may want to have a clean sample with little contamination or a sample that captures the largest number of objects at the expense of a larger contamination level. In contrast to many other classification tasks, we have a good estimate of the correct number of stars—just not which *individual* eROSITA sources are the stars. Therefore, we chose to optimize the classifier such that the correct number of sources are classified as stars, that is, have at least one likely stellar association. This choice implies that the number of stars misclassified as non-stellar N_{missed} and the number of sources erroneously classified as stars $N_{spurious}$ are equal. However, this also implies that we do not necessarily achieve the highest possible accuracy; it is conceivable that a solution exists with a larger number of correctly classified objects at the cost of, e.g., a larger imbalance between N_{missed} and $N_{spurious}$. Inspecting the behavior of the SVC, we found that the hyper-parameters resulting in $N_{missed} = N_{spurious}$ span a well-defined path in the parameter space. From those solutions, we chose the parameters that achieve the highest accuracy.

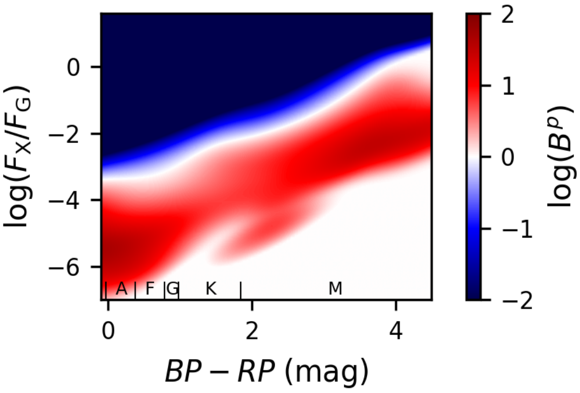


Fig. 5. Distribution of the Bayes factor B^p as a function of the activity F_X/F_G and the color $BP - RP$ for the eFEDS counterparts. The ranges of the spectral types are indicated at the bottom of the figure.

3.4. SVC results

We used a third degree polynomial kernel and the properties from Tab. 1 replacing the X-ray and optical fluxes with their flux ratio F_x/F_G resulting in a six dimensional feature vector. The SVC was then trained with the above described training sample and the hyper-parameters adjusted so that $N = 2060$ eFEDS sources are classified as stellar. With this requirement, the expected number of eROSITA sources randomly classified as stellar within the 2060 sources is $N_{spurious} = 239$ so that also $N_{missed} = 239$ real sources are not classified as stellar by the SVC. This corresponds to a completeness and reliability of 88.4 % (cf. Eqs. 5 and 6 for the definitions of completeness and reliability). The resulting sample properties will be discussed in sect. 5 together with the sample resulting from the Bayesian approach.

The SVC does not directly provide well calibrated probabilities p_{ij} for the associations $i \leftrightarrow j$. Therefore, one cannot easily construct samples with different completeness or reliability values applying certain cuts in association probability p_{ij}; rather a new training run would be required to achieve the best performance for a specific completeness or reliability level. However, the number of random associations for a certain cutoff in p_{ij} can be directly derived by applying the algorithm to the validation sample. A natural choice is therefore to empirically calibrate the

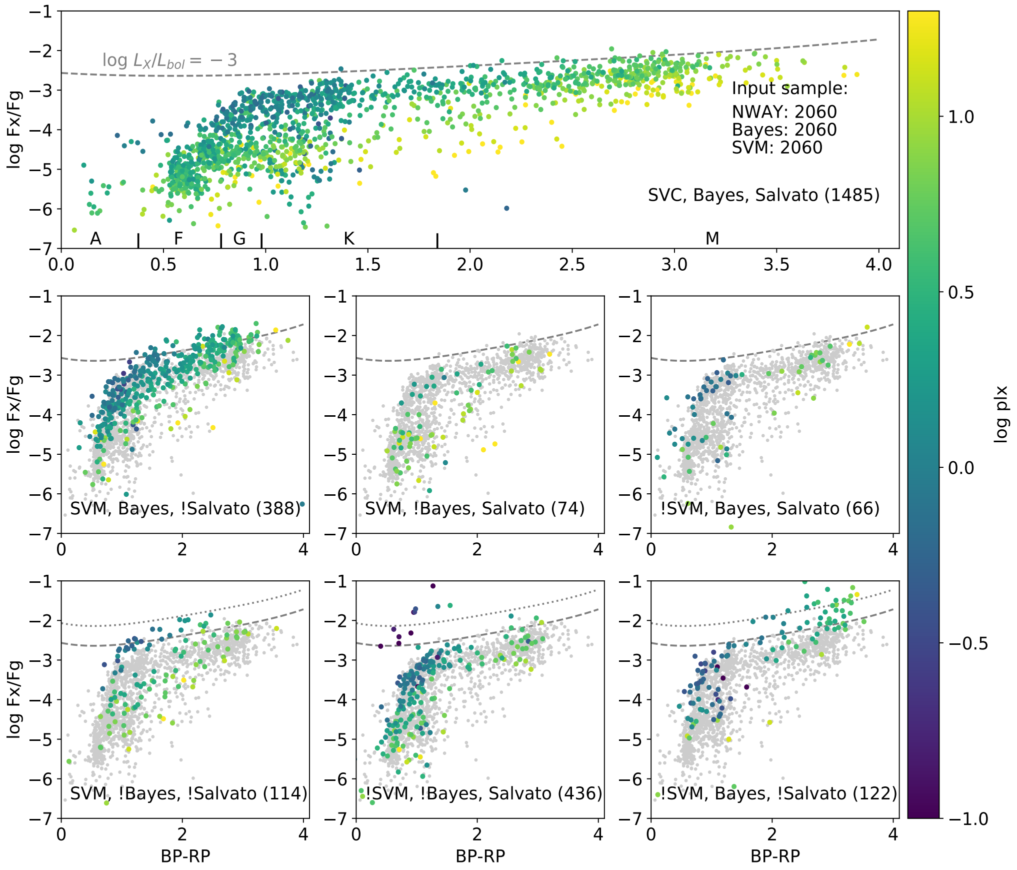


Fig. 6. Ratio between X-ray and G-band fluxes as a function of the associated Gaia object's BP-RP color. The objects are colored according to the parallax. The label in the lower part of all seven panels indicate in which sample the colored objects belong: SVM, Bayes, and Salvato describe the method and "!" equals *not*, that is, that the colored identification are not identified by the method that is preceded by the "!". The number in bracket indicates the respective number of sources. In rows two and three, the gray dots represent the objects of the top panel for reference. The dotted line in the bottom row indicates $\log L_X/L_{bol} = -2.5$.

association probabilities p_{ij} such that

$$N_{spurious}(p > p_{min}) = \sum_{p>p_{min}} 1 - p_{ij} \tag{8}$$

using the validation sample. Figure 4 shows the resulting completeness and reliability levels as a function of the cutoff value p_{min} in the thus calibrated association probabilities p_{ij}. Here, we used the previous definitions for completeness and reliability replacing N with $N_>(p_{min})$, i.e., the number of associations above the cutoff-threshold.

4. Method II: Bayesian approach

In our Bayesian matching framework, the prior probability of picking by chance the correct counterpart is updated after obtaining data of the source position and properties. In that sense, we followed similar Bayesian catalog matching techniques described by Budavári & Szalay (2008) and successfully implemented in cross-matching tools such as NWAY (Salvato et al. 2018).

4.1. Distance based matching probability

Again, we consider the problem of matching N_G eligible stellar candidates to N_X X-ray sources. Therefore, we define the following hypotheses, the probabilities of which we want to compare:

H_{i1} : the i-th X-ray source is associated with the $j = 1$ counterpart

$\vdots$

H_{ij} : the i-th X-ray source is associated with the j-th counterpart

H_{i0} : the i-th X-ray source is not associated with any of the counterparts.

Extending upon the scheme of Budavári & Szalay (2008), we know that only a fraction CF of the X-ray sources can be associated with one of the N_G counterparts based on the distribution of the match distances r_{ij} and sky densities η_j (cf. Sect. 2.2). Hence, we derived the prior probability that none of the optical counterparts is associated with the i-th X-ray source through

$$P(H_{i0}) = 1 - \mathrm{CF}, \tag{9}$$

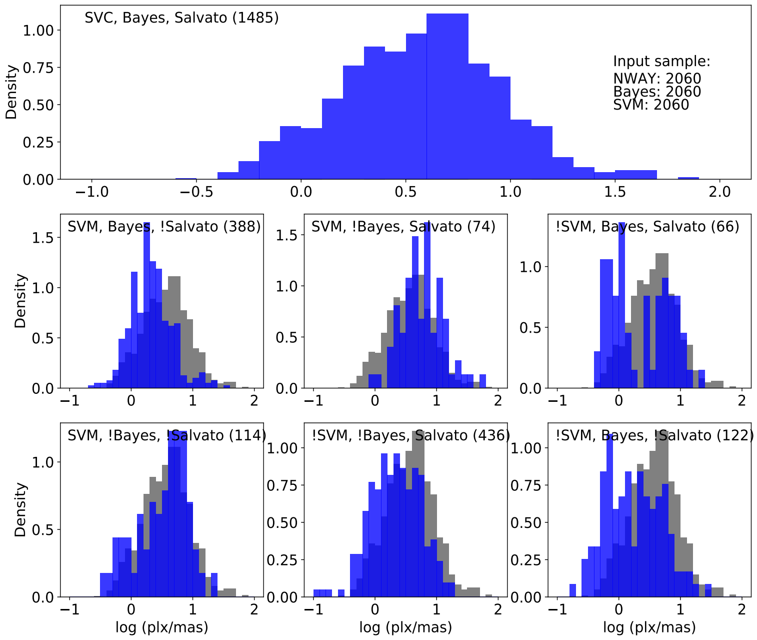


Fig. 7. Histograms for the parallaxes of the identified Gaia counterparts. Individual panels as in Fig. 6. The parallax distribution of the associations identified by all three methods (top panel) is shown in the background as the gray histogram in the middle and bottom rows.

and the prior probability that the j-th counterpart is the identification of the i-th X-ray source by

$$P(H_{ij}) = \frac{\mathrm{CF}}{N_G} = \frac{\mathrm{CF}}{\eta\Omega}, \tag{10}$$

where η is the source density in the area of the sky given by Ω.

Next, we considered the positions of the X-ray source and the eligible stellar candidates. Neglecting the positional uncertainty of the stellar candidates, the likelihood of obtaining the data D_i given that the j-th counterpart is the correct identification of the i-th X-ray source was estimated through

$$P(D_i|H_{ij}) = \frac{1}{2\pi\sigma_i^2} e^{-\frac{r_{ij}^2}{2\sigma_i^2}} , \tag{11}$$

where σ_i is the positional uncertainty of the i-th X-ray source and r_{ij} is the angular separation between the i-th X-ray source and the j-th counterpart (cf. Eq. 1). Counterparts for which $r_{ij} \gg \sigma_i$ can be neglected based on the exponential term. Assuming the X-ray source is not associated with any of the counterparts, the likelihood of the data becomes

$$P(D_i|H_{i0}) = \frac{1}{4\pi} . \tag{12}$$

In accordance with Budavári & Szalay (2008), we then obtained the geometric Bayes factor

$$B_{ij}^g = \frac{P(D_i|H_{ij})}{P(D_i|H_{i0})} = \frac{2}{\sigma_i^2} e^{-\frac{r_{ij}^2}{2\sigma_i^2}} . \tag{13}$$

Applying Bayes' theorem, we computed the posterior probability that the j-th counterpart is the correct identification through

$$p_{ij} = P(H_{ij}|D_i) = \frac{P(D_i|H_{ij}) \cdot P(H_{ij})}{\sum_{k=0}^{N_G} P(D_i|H_{ik}) \cdot P(H_{ik})} . \tag{14}$$

The probability that any of the stellar counterparts is the correct identification, which is equivalent to the X-ray source being stellar given a complete and uncontaminated counterpart catalog of all and only stars, is given by

$$p_{\mathrm{stellar}} = \frac{\sum_{k=1}^{N_G} P(D_i|H_{ik}) \cdot P(H_{ik})}{\sum_{k=0}^{N_G} P(D_i|H_{ik}) \cdot P(H_{ik})}. \tag{15}$$

4.2. Consideration of additional source properties

Additional properties of the counterparts and X-ray sources can be considered to identify the best match. For example, the matching probability can be increased if the activity estimated by the X-ray flux and the optical brightness of the counterpart meets the expectations of a stellar source and few random associations are expected at such activity levels. Technically, this was achieved by expanding the geometric Bayes factor by another factor B_{ij}^p, which represents the expected ratio between physical associations and random associations, so that

$$B_{ij} = B_{ij}^g \times B_{ij}^p . \tag{16}$$

As additional properties, we used the X-ray to optical flux ratio, F_X/F_G, and the $BP - RP$ color.

We estimated the factor B_{ij}^p from the data themselves. In particular, we constructed a training set of highly probable geometric counterparts. To obtain a clean training set, we selected the counterparts with a posterior geometric matching probability > 0.9 only (cf. Eq. 14). Of the thus identified 577 reliable counterparts, we expect 32 to still be spurious. Applying the same screening procedures as for the SVC training sample (cf. Sect. 3.1), we arrived at a training set with 494 bona fide coronal sources. To derive the distribution of specific properties for spurious association, we again shifted the X-ray sources arbitrarily on the fixed background of counterparts in the eFEDS field and studied the thus constructed set of explicitly random matches.

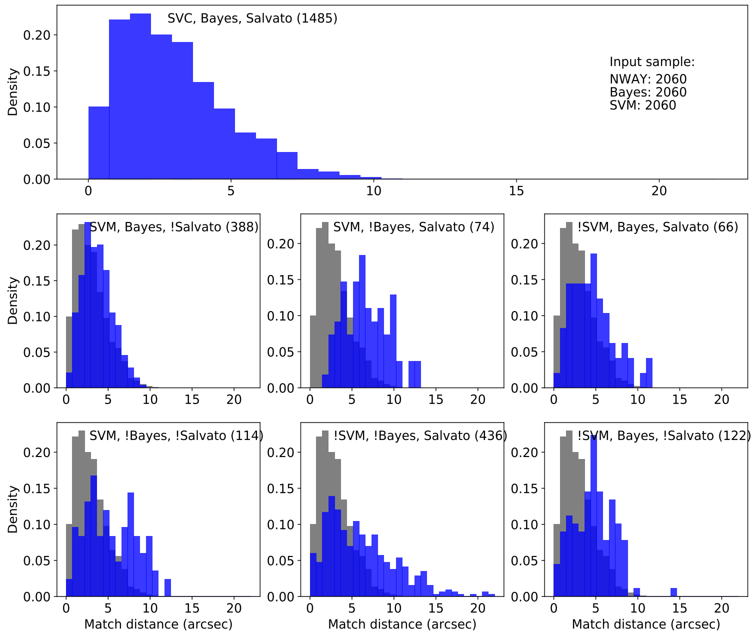


Fig. 8. Histograms for the match distance to the associated Gaia counterparts. Similar to Fig. 7.

As our training sample remains rather small, some regions in the F_X/F_G versus $BP - RP$ plane remain sparsely populated, which confounds the estimation of B^p_{ij}. This affects, e.g., M-type sources at low F_X/F_G values, which are reasonable X-ray emitters by physical standards, but are unlikely to be detected in X-rays at the eFEDS sensitivity. Therefore, we adopted a value of one for B^p_{ij} in this region of inactive M-dwarfs. On the other end of the main sequence, early A- and B-type stars with high F_X/F_G values can be excluded as true identifications based on physical grounds, here B^p_{ij} goes to zero, which is tantamount to assuming that essentially all such matches are random associations. Nonetheless, for most of the counterparts, the details of the estimation of B^p_{ij} have minor impact on the derived matching probabilities.

In Fig. 5 we show the resulting map of Bayes factors B^p_{ij} for the eFEDS counterparts. For example, F-type sources with an activity level of around $\log(F_X/F_G) \approx -5$ are weighted up, while counterparts with $\log(F_X/F_G) > -2$ are weighted down by considering the additional properties. The Bayes map generally appears rather smooth. However, regions of the Bayes map sparsely populated by training and control sources may show some distinct structure, in particular the increase of the Bayes factor for sources around $F_X/F_G = -5$ and $BP - RP = 2.2$ mag. The influence of these "low source density regions" on our identifications is marginal because the number of counterparts in these regions is also very small and in fact, neither the SVM nor the Salvato et al. method show an excess or deficit of sources in this part of the diagram when compared to the Bayesian method (see Fig. 6). We also note that although sources with high X-ray to optical flux ratios are excluded from the training set, such sources can nevertheless be identified in the final sample if their positional match is good or the number of expected spurious association with such properties is small because then the Bayes factor B^p_{ij} is still about unity.

4.3. Bayesian results

Applying the matching procedure described in Sects. 4.1 and 4.2, we obtained a stellar probability for every eFEDS source. In contrast to the SVC, these probabilities are well calibrated so that the completeness and reliability of a sample selected by a specific probability threshold can be derived directly. The expected number of missed (false negatives) and spurious (false positives) stellar identification is estimated through

$$N_{\rm missed} = \sum_{N_<} p_{\rm stellar,<}, \tag{17}$$

$$N_{\rm spurious} = \sum_{N_>} 1 - p_{\rm stellar,>} \tag{18}$$

where $N_>$ and $N_<$ are the number of sources above and below the threshold and their probabilities are denoted by $p_{\rm stellar,>}$ and $p_{\rm stellar,<}$, respectively. Completeness and reliability of the obtained sample were estimated through Eqs. 5 and 6 by replacing N with $N_>$.

In Fig. 4 we present the expected completeness and reliability obtained with the Bayesian approach as a function of the stellar probability cutoff. At $p_{\rm stellar} \approx 0.58$ the expected number of stellar sources in the eFEDS field is recovered, here, about 11 % of the identifications are expected to be spurious and the same fraction of stellar eFEDS sources is expected to be missed. Going to larger stellar probabilities, the completeness decreases and the reliability increases, as expected. We note that these values were empirically verified by applying our Bayesian identification procedure to arbitrarily shifted eFEDS sources similar to the SVC test sample.

5. Results and comparison between approaches

We are interested in the physical properties of the stars detected by eROSITA, more precisely, in the resulting association properties. Therefore, we base our comparison on associations and not

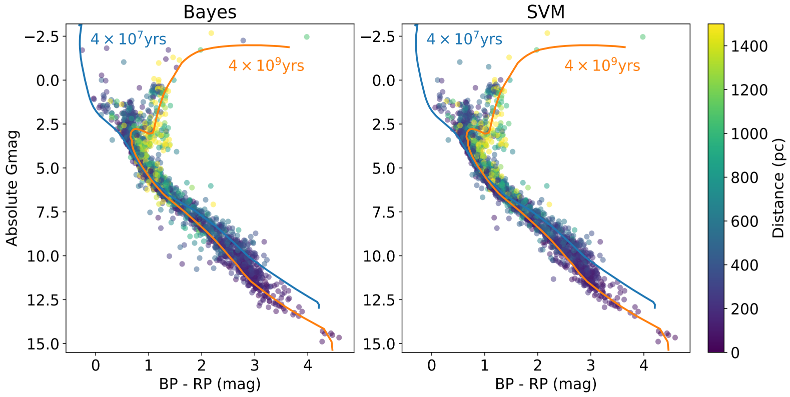


Fig. 9. Color-magnitude diagrams for the identified stellar sources (**Left**: Bayes, **right**: SVM). The color indicates the distance of the sources. Isochrones are shown for two representative stellar ages of 4×10^7 and 4×10^9 years using the PARSEC isochrones (Bressan et al. 2012).

on the classification of an eROSITA source as stellar; a discussion of the overlap in stellar classifications is provided in Salvato et al. (2021).

5.1. Comparison samples

To compare the samples identified by the different methods, we cut the SVM and Bayesian catalogs in p_{ij} such that 2060 eROSITA sources are classified as stellar. We also include the catalog presented in Salvato et al. (2021), which provides counterparts to all eFEDS sources regardless of their galactic or extra galactic nature. These authors used a training sample of 23 000 *XMM-Newton* and *Chandra* sources with secure counterpats and the final counterpart identification is done after comparing the associations obtained with NWAY (Salvato et al. 2018) with a modified version of Maximum Likelihood (Sutherland & Saunders 1992) as described in Ruiz et al. (2018). Again, we cut their sample of stellar sources in p_{any} to obtain 2060 stellar eROSITA sources. Furthermore, we mapped their Gaia DR2 IDs to EDR3 IDs for the comparison noting that not all DR2 sources could be unambiguously mapped to a EDR3 source.

The SVM and Bayesian approaches perform about equally well with an overlap of 1873 associations in both samples (91 %). This fraction incidentally coincides with the expected reliability of the respective samples. Therefore, it is possible albeit unlikely that all true associations are among the shared associations and that all spurious associations are being identified by one method alone. The resulting sample properties in terms of completeness and reliability are very similar for the SVM and Bayesian approaches (see Fig. 4). The differences are compatible with the Poisson noise for the number of missed or spurious sources at the respective $p_{stallar}$-thresholds.

The overlap between the results by Salvato et al. (2021) and the SVM and Bayesian approaches is smaller with about 1550 and 1558 associations in common, respectively. This smaller overlap is partly caused by the different Gaia data releases (145 associations are affected) and by the construction of the sample since we cut the sample to the 2060 “best” associations. The overlap in stellar classifications of eROSITA sources is much higher (cf. Salvato et al. 2021).

5.2. Sample properties

We show the physical properties of the resulting samples in Figs 6 to 9. Figures 6 to 8 display different properties but their structure is identical. They all contain seven panels organized in three rows: The first row shows the properties of the associations overlapping between all three methods (1484 in total). The second row shows the properties of the associations overlapping between only two methods and we present one panel for each of the possible combinations. For each panel, we provide an annotation indicating which sample(s) contain the displayed associations. We use a “`!`” to indicate that the associations are not in the sample following the “`!`”, for example, the left panel in the second row with the `SVM, Bayes, !Salvato` annotation implies that the associations are in the SVM and Bayesian samples but not among the 2060 best Salvato et al. (2021) stellar associations. Finally, the third row shows the associations exclusively contained in only one sample. To ease the comparison between the different samples, we display in rows two and three of Figs. 6 to 8 the associations overlapping between all three methods (shown in the top panel of each figure) in the background (in gray). Numbers in brackets following the sample description indicate the number of associations (displayed in color).

Figure 6 shows that the F_X/F_G distributions are relatively similar for all identification methods Noticeable is that all methods classify some associations above the saturation limit of $\log L_X/L_{bol} = -3$ as stellar (gray dashed line in Fig. 6). These associations between an eROSITA source and a Gaia source have a combination of high positional match likelihoods and are, for SVC, relatively nearby, which counterbalances the high F_X/F_G-values and lead to the stellar classification. At least some cases of unusually high F_X/F_G values result from flaring: 65 variable sources are discussed in Boller et al., this issue. Most of these variable sources are associated with stars and show, when sufficient X-ray coverage is available, that the X-ray light curves resemble the typical shape of flaring due to magnetic activity. Furthermore, most sources are within 0.5 dex of the boundary for all methods (dashed line in Fig. 6). The methods employed by Salvato et al. (2021) identify a small number of sources (~ 10) with Gaia sources leading to quite high F_X/F_G values, probably because their association is based on charaterizing the entire SED. In total, the number of sources above the saturation limit remains small: neither method associates more than hundred eFEDS sources with a Gaia source leading to F_X/F_G values implying X-ray emission above the saturation limit (< 5 % of the stellar classifications).

Overall the associations classified as stellar fall within the expected BP-RP-color vs F_X/F_G range of stellar sources for all three methods. The main features are (a) high X-ray activity levels for M dwarfs close to the saturation level, (b) a substantial spread in X-ray activity for stars of spectral types F to early K, and (c) the onset of magnetic activity at late A/early F spectral types. These properties are also seen for the sample of stellar X-ray sources in the XMM-Newton slew survey catalog identified by Freund et al. (2018) and may be a general property of flux limited X-ray surveys.

We show in Fig. 9 the positions of the Gaia associated with the stellar eROSITA in the color-magnitude diagram. By construction, the associated Gaia sources occupy positions compatible with young to main sequence stars as well as stars of moderate age (in the few Gyr range). Differences between the SVM and Bayesian method are small and the identified Gaia sources are largely overlapping in the HRD. As expected, the distances towards the earlier spectral types tend to be larger than towards the later spectral types, in particular M dwarfs, which are found within about 200 pc. Both methods associated stars on the red giant branch with eROSITA sources. These sources may not be causing the X-ray emission as red giants beyond a so-called dividing line have been found to lack genuine X-ray emission (Haisch et al. 1991). We note, however, that (a) lower-mass companions may be responsible for the detected X-ray emission and (b) the association likelihoods of these sources are indeed quite low. In particular, the reddest giant in Fig. 9 (HT Hya) is also a GALEX FUV and NUV source with $p_{stellar} \approx 0.7$, i.e., this association may be very well be indeed correct although the giant itself may not be the source of the X-ray emission. Nevertheless, we purposely keep these sources (and potentially other source classes) in the final sample to avoid biasing ourselves unduly towards current "wisdom" precluding new discoveries.

The parallax-distributions (Fig. 7) show that the identified stellar sources peak between log plx=0.5 and log plx=1, i.e., are mostly between 100 pc and 250 pc from the Sun. Few sources are farther than 1 kpc (log plx<0). The objects associated by the Bayesian method and Salvato et al. (2021) tend to have smaller parallaxes (larger distances) compared to the SVC. For the Bayesian method, this may be expected, because it is currently ignorant of the Gaia parallaxes. Salvato et al. (2021) do consider the parallax in their the method, but it likely has a different (lower) importance than for the SVC, which can use the parallax information to boost the association likelihood of certain associations (see red dots in Fig. 10). In theory, the parallax information may be beneficially used to reduce the number of spurious associations as the number of possible random associations increases with distance. At the moment, however, the SVC is not able to take full advantage of this benefit resulting in similar reliability and completeness levels as the Bayesian approach, probably because of the limited training set.

Lastly, the match distance-distribution shows that the associations identified as stellar by all three methods have, on average, the smallest match distances (cf. Fig. 8). The stellar associations shared between the Bayesian and SVC methods have relatively small match distances, too. The other panels of Fig. 8, however, show that the match distances tend to be larger for associations found by only one method. For these associations, the positional match is insufficient to result in a secure stellar identification and the weighting of the other features becomes important, which is different between the three methods.

5.3. Probabilities

Our Bayesian and SVC methods were developed focusing on stars (coronal emitters). Therefore, we compare the calculated probabilities from the SVC and Bayesian methods in Fig. 10 (after calibrating the SVC probabilities as described in sect. 3.4). Overall, the probabilities are very similar being clustered around the 1:1 relation, most sources get either very low or high association probabilities in both methods. In addition to some intrinsic scatter, a number of sources get relatively high association probabilities by the SVC method while the Bayesian method assigns only mediocre probabilities between 0.5 and 0.8 (the structure located in the right middle part of Fig. 10. Most of these sources are within 100 pc (red dots in Fig. 10) so that they get boosted association probabilities due to these high parallax-values compared to the Bayesian approach, which currently ignores the parallaxes.

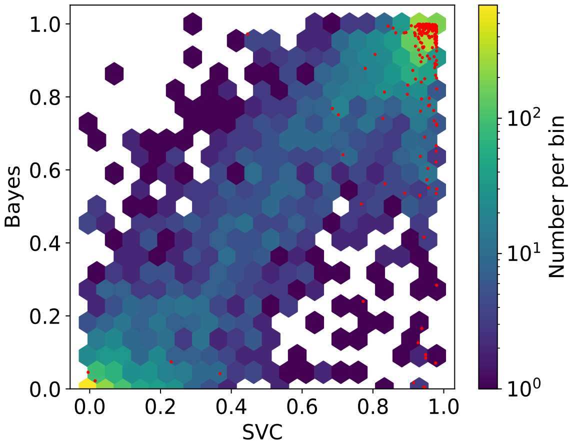


Fig. 10. Association probabilities for the SVC and Bayesian methods. The red dots depict the sources within 100 pc.

5.4. HamStar Catalog

The catalog of the stellar sources is based on the Bayesian framework, because it directly provides probabilities reflecting our prior knowledge and the data. We also include two columns that indicate if the particular association is also identifified by the SVM approach and by NWAY (Salvato et al. 2021). This catalog is dubbed "HamStars" in the eROSITA context.

6. Conclusions and Outlook

We present SVC and Bayesian methods specifically designed to identify stars among the eROSITA sources. Both methods provide very similar results, both, in terms of sample quality (completeness and reliability) and resulting physical properties. Both methods provide significant improvements over a purely geometric approach reducing the number of spurious associations in the sample by about a factor of two and samples constructed to contain the geometrically expected number of eROSITA sources achieve almost 90 % completeness and reliability. Furthermore, we show how to construct calibrated probabilities for both methods, which can be eventually used to create sub-samples with specific completeness and reliability properties by applying appropriate cuts in association probability.

On the one hand, it is somewhat surprising that the SVC and Bayesian methods perform so similarly, because the SVC method uses the parallax measurement as an additional information compared to the Bayesian approach. On the other hand, the SVC method is ignorant of the "mathematics" of the matching distance and needs to learn the positional match characteristics from the training sample. We expect that a larger training sample will, at least partly, mitigate this issue, in particular when applied to the eROSITA all-sky survey (eRASS).

The sample quality in terms of completeness and reliability is specific to eFEDS, because these properties depend on (a) the depth of the X-ray exposure, (b) the ratio between stellar and other X-ray sources, and (c) the sky density of the eligible stellar counterparts. These properties are expected to change strongly with galactic position, e.g. the sky density of eligible stellar counterparts differs by at least a factor of hundred between the galactic pole and bulge regions. Therefore, these effects need to be taken into account to achieve similar or better results for eRASS compared to eFEDS. In addition, In the future, we plan to also include additional information from the X-ray data such as spectral hardness to further improve the algorithms for their application to eRASS and it is straight forward to include additional information in both methods. Lastly, future Gaia data releases will improve the quality of the match catalog. We are therefore confident that it is possible to construct well characterized stellar samples from eROSITA data. This large sample of stars with well known X-ray properties will allow us to improve our understanding of stellar activity throughout space and time.

Acknowledgements. eROSITA is the primary instrument aboard SRG, a joint Russian-German science mission supported by the Russian Space Agency (Roskosmos), in the interests of the Russian Academy of Sciences represented by its Space Research Institute (IKI), and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The SRG spacecraft was built by Lavochkin Association (NPOL) and its subcontractors, and is operated by NPOL with support from IKI and the Max Planck Institute for Extraterrestrial Physics (MPE). The development and construction of the eROSITA X-ray instrument was led by MPE, with contributions from the Dr. Karl Remeis Observatory Bamberg & ECAP (FAU Erlangen-Nürnberg), the University of Hamburg Observatory, the Leibniz Institute for Astrophysics Potsdam (AIP), and the Institute for Astronomyand Astrophysics of the University of Tübingen, with the support of DLR and the Max Planck Society. The Argelander Institute for Astronomy of the University of Bonn and the Ludwig Maximilians Universität Munich also participated in the science preparation for eROSITA. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. PCS gratefully acknowledges support by the DLR under 50 OR 1901 and 50 OR 2102. SF acknowledge supports through the Integrationsamt Hildesheim, the ZAV of Bundesagentur für Arbeit, and the Hamburg University and thanks Gabriele Uth and Maria Theresa Lehmann for their support.

References

Binder, D. A. 1978, Biometrika, 65, 31
Boller et al., T. 2021, A&A
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Brunner et al., H. 2021, A&A
Budavári, T. & Szalay, A. S. 2008, ApJ, 679, 301
Cortes, C. & Vapnik, V. 1995, Mach Learn, 20, 273–297
Fischer, R. A. 1938, Annals of Eugenics, 8, 376
Freund, S., Robrade, J., Schneider, P. C., & Schmitt, J. H. M. M. 2018, A&A, 614, A125
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2020, arXiv e-prints, arXiv:2012.01533
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
Güdel, M. 2004, A&A Rev., 12, 71
Haisch, B., Schmitt, J. H. M. M., & Rosso, C. 1991, ApJ, 383, L15
Marton, G., Ábrahám, P., Szegedi-Elek, E., et al. 2019, MNRAS, 487, 2522
Melton, E. 2020, AJ, 159, 200
Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv e-prints, arXiv:1209.3114
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., & Ventura, P. 2003, A&A, 397, 147
Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
Ruiz, A., Corral, A., Mountrichas, G., & Georgantopoulos, I. 2018, A&A, 618, A52
Salvato, M., Buchner, J., Budavári, T., et al. 2018, MNRAS, 473, 4937
Salvato et al., M. 2021, A&A
Schmitt, J. H. M. M. & Liefke, C. 2004, A&A, 417, 651
Skumanich, A. 1972, ApJ, 171, 565
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Sutherland, W. & Saunders, W. 1992, MNRAS, 259, 413
Testa, P., Saar, S. H., & Drake, J. J. 2015, Philosophical Transactions of the Royal Society of London Series A, 373, 20140259
Vilhu, O. 1984, A&A, 133, 117
Vioque, M., Oudmaijer, R. D., Schreiner, M., et al. 2020, A&A, 638, A21
Wright, N. J., Drake, J. J., Mamajek, E. E., & Henry, G. W. 2011, ApJ, 743, 48

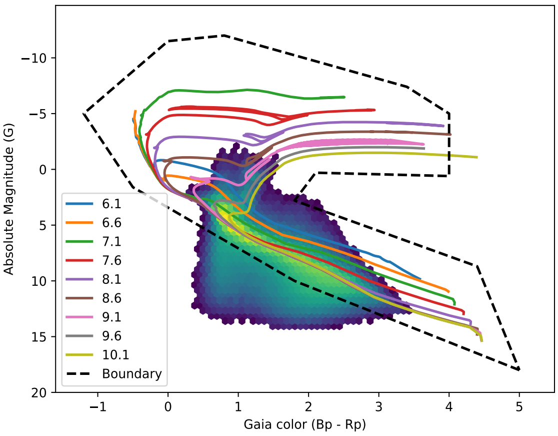


Fig. A.1. Polygon containing valid stellar sources with PARSEC evolutionary tracks. The 2D histogram in the background shows a typical Gaia source population; most sources are close to the 100+ Myrs main sequence.

Appendix A: Isochrones

Figure A.1 shows the region in which a Gaia source is considered an eligible stellar counterpart.

5.2 Publication: The eROSITA Final Equatorial-Depth Survey (eFEDS): Identification and characterization of the counterpart to the point-like sources

Astronomy & Astrophysics manuscript no. aanda
June 29, 2021

The eROSITA Final Equatorial-Depth Survey (eFEDS):

Identification and characterization of the counterparts to the point-like sources

M. Salvato⋆[1,2], J. Wolf[1,2], T. Dwelly[1], A. Georgakakis[3], M. Brusa[4,5], A. Merloni[1], T. Liu[1], Y. Toba[6,7,8], K. Nandra[1], G. Lamer[9], J. Buchner[1], C. Schneider[10], S. Freund[10], A. Rau[1], A. Schwope[9], A. Nishizawa[11], M. Klein[12], R. Arcodia[1], J. Comparat[1], B. Musiimenta[4,5], T. Nagao[8], H. Brunner[1], A. Malyali[1], A. Finoguenov[1], S. Anderson[13], Y. Shen[14], H. Ibarra-Mendel[14], J. Trump[15], W.N., Brandt[16], C.M., Urry[17], C. Rivera[7], M. Krumpe[7], T. Urrutia[9], T. Miyaji[18], K. Ichikawa[1,19,20], D.P., Schneider[21,22], A. Fresco[1], J. Wilms[23] T. Boller[1], J. Haase[1], J. Brownstein[24], R.R., Lane[25], D. Bizyaev[26], C. Nitschelm[27]

(Affiliations can be found after the references)

Received September XX, 2020; accepted XX, 2020

ABSTRACT

Context. In Fall 2019, eROSITA on board of SRG observatory started to map the entire sky in X-rays. After the 4-year survey program, it will reach flux limits about 25 times deeper than ROSAT. During the SRG Performance Verification phase, eROSITA observed a contiguous 140 deg^2 area of the sky down to the final depth of the eROSITA all-sky survey ("eROSITA Final Equatorial-Depth Survey": eFEDS), with the goal of getting a census of the X-ray emitting populations (stars, compact objects, galaxies, clusters of galaxies, AGN) that will be discovered over the entire sky.

Aims. This paper presents the identification of the counterparts to the point-sources detected in eFEDS in the Main and Hard samples described in Brunner et al., and their multi-wavelength properties, including redshift.

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

Results. The eFEDS sources with a reliable counterparts are 24774/27369 (90.5%) in the Main sample and 231/246 (93.9%) in the Hard sample, including 2514 (3) sources for which a second counterpart is equally likely. By means of reliable spectra, Gaia parallaxes, and/or multiwavelength properties we have classified the counterparts in both samples as 'Galactic' (2822) and 'extragalactic' (21952). For about 340 of the extragalactic sources we cannot rule out the possibility that they are unresolved clusters or belong to clusters. Inspection of the distributions of the X-ray sources in various optical/IR color-magnitude spaces reveal a rich variety of diverse classes of objects. The photometric redshifts are most reliable within the KiDS/VIKING area, where also deep near-infrared data is available.

Conclusions. This paper is accompanying the eROSITA early data release of all the observations performed during the performance and verification phase. Together with the catalogs of primary and secondary counterparts to the Main and Hard samples of the eFEDS survey this paper releases their multiwavelength properties and redshifts.

Key words. quasars: individual – Galaxies: high-redshift – X-rays: galaxies

1. Introduction

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

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

eROSITA (extended ROentgen Survey with an Imaging Telescope Array; Predehl et al. 2021), onboard the Spektrum-Roentgen-Gamma (SRG) mission, was designed to provide sensitive X-ray imaging and spectroscopy over a large field of view, thus unlocking unprecedented capabilities for surveying large areas of the sky to deep flux levels. Moreover, the SRG mission plan includes a long (4 years), uninterrupted all-sky survey program (the eROSITA All-Sky Survey: eRASS; Predehl et al. 2021) capable of detecting, for the first time, millions of X-ray sources.

In order to demonstrate these ground-breaking survey capabilities, and prepare for the science exploitation of the upcoming all-sky survey, the contiguous 140 square degrees of the eROSITA Final Equatorial-Depth survey (eFEDS; Brunner et al., submitted) was observed during the SRG Calibration and Performance Verification phase, between the 3rd and the 7th of

---

⋆ mara@mpe.mpg.de

November 2019. The entire field was observed to an approximate depth of ~2.2 ks (~1.2 ks after correcting for telescope vignetting), corresponding to a limiting flux of $F_{0.5-2\,\mathrm{keV}} \sim 7\times10^{-15}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$. The eFEDS field was chosen from among the extragalactic areas with the richest multi-wavelength coverage visible by eROSITA in Fall 2019. The observations are just about 50% deeper than anticipated for eRASS:8 at the end of the planned 4-year' program in the ecliptic equatorial region ($\sim 1.1\times10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$, Predehl et al. 2021). As such, eFEDS is a fair representation of what the final eROSITA all-sky survey will be, enabling scientists to face and solve the challenges that will accompany their work for the duration of the survey.

As discussed in detail in Brunner et al. (submitted), the X-ray catalogs generated by the analysis of the eFEDS eROSITA data comprise a *Main* one, with 27910 sources detected above a detection likelihood of 6 in the most sensitive 0.2-2.3 keV band, and a *Hard* one, containing 246 sources detected above a detection likelihood of 10 in the less-sensitive 2.3-5 keV band.

In this paper, we focus our attention on the point-like (i.e. with an extension likelihood `EXT_LIKE=0`[1]) X-ray sources contained in these catalogs (27369 and 246 for the *Main* and *Hard* sample, respectively), and describe in detail the procedure to (i) identify reliably multi-wavelength counterparts to the eROSITA sources, (ii) classify and characterize their properties and (iii) provide reliable redshift measurements (spectroscopic when available and photometric otherwise). The identification and determination of the reliability of the counterparts, the computation of the photometric redshifts (photo-z), and the characterisation of the sample follow the same procedure for both *Main* and *Hard* samples, and for simplicity we discuss here specifically only the *Main* sample, given the large overlap between the two catalogs (226/246 hard sources are in common). While we provide here the catalog of counterparts for all the sources in both samples, the properties of the sources in the *Hard* sample are presented and discussed in Nandra et al. (in prep.). The papers on X-ray spectral analysis (Liu, T. et al., submitted), variability (Boller et al., Buchner et al., submitted), X-ray Luminosity Function (Buchner et al., in prep.; Wolf et al., in prep.) and host properties of eFEDS AGN (Li et al., in prep.) are all based on the catalog and/or the methodology presented in this work. Based on this work are also the papers presenting interesting single objects (Brusa et al., submitted, Toba et al. 2021, Wolf et al. 2021), the X-ray properties of WISE sources in eFEDS (Toba et al. submitted) and photo-z computed via machine learning (Nishizawa et al., in prep),

The structure of the paper is as follows: In section 2 we summarise the availability of ancillary data that will be used for the identification of the X-ray counterparts and photo-z estimates. Section 3 describes in detail the method used in this paper for identifying the counterparts. Because of its size, the field is well populated by stars, AGN, clusters and nearby galaxies. Each eROSITA working group has developed independent methods for the identification of sources of interest, and in Section 4 a comparison is made with two main source classes: stellar coronal emitters and clusters of galaxies, with the ultimate goal of consolidating the counterparts and classifying them at the same time. Section 6 presents and discusses the photo-z computed with Le PHARE (Ilbert et al. 2006; Arnouts et al. 1999), including a comparison with DNNz (Nishizawa et al., in prep), an independent method based on machine learning. Section 7 describes the released data. The basic properties of the point-source eFEDS population based on redshift, photometry, and X-ray flux are presented in Section 8. The conclusions in Section 9 close the paper, with a forecast of the results and challenges that we will face when working on the eROSITA all-sky survey.

The description of the Catalogs that we release is provided in the Appendix, together with the list of templates used for computing the photo-z.

Throughout the paper we assume AB magnitudes unless differently stated. In order to allow direct comparison with existing works from the literature of X-ray surveys, we adopt a flat ΛCDM cosmology with $h = H_0/[100\,\mathrm{km\,s^{-1}Mpc^{-1}}] = 0.7$; Ω_M=0.3; Ω_Λ=0.7.

2. Supporting ancillary data

For studies of X-ray sources (taken singularly or as a population) the entire spectral energy distribution (SED) needs to be constructed and the redshift determined. Only rarely can redshift be obtained directly from X-ray spectra, and it is instead routinely obtained either via optical/Near-Infrared spectroscopy or via photometric techniques. However, for that to work the counterparts to the X-ray sources need to be determined first. Deep and homogeneous multi-wavelength data are therefore a pre-requisite for any complete population study of an X-ray survey.

The main challenge is that, in survey mode, eROSITA has a Half Energy Width (HEW) of 26 arcsec[2] (Predehl et al. 2021), which makes the identification of the correct counterparts not at all trivial (e.g., see the review of Naylor et al. 2013; Salvato et al. 2018b), especially if considering in addition that wide field multi-wavelength homogeneous surveys are very difficult to obtain, with very few exceptions (see the Legacy imaging survey supporting the Dark Energy Spectroscopic Instrument; Dey et al. 2019a). By construction, the eFEDS field is placed in an area fully encompassing the GAMA09 equatorial field (Driver et al. 2009), which has deep optical/NIR imaging thanks to the HSC Wide area Survey (Aihara et al. 2018a), KiDS/VIKING (Kuijken et al. 2019a), DESI Legacy Imaging Survey (Dey et al. 2019b), and, among others, GAMA (Driver et al. 2009), WIGGLEz (Drinkwater et al. 2018a), LAMOST[3] and SDSS (Blanton et al. 2017) spectroscopic coverage. Below we list and describe in more detail the surveys that have been used in this work. Table 1 summarises the depth in each filter.

2.1. Supporting the associations

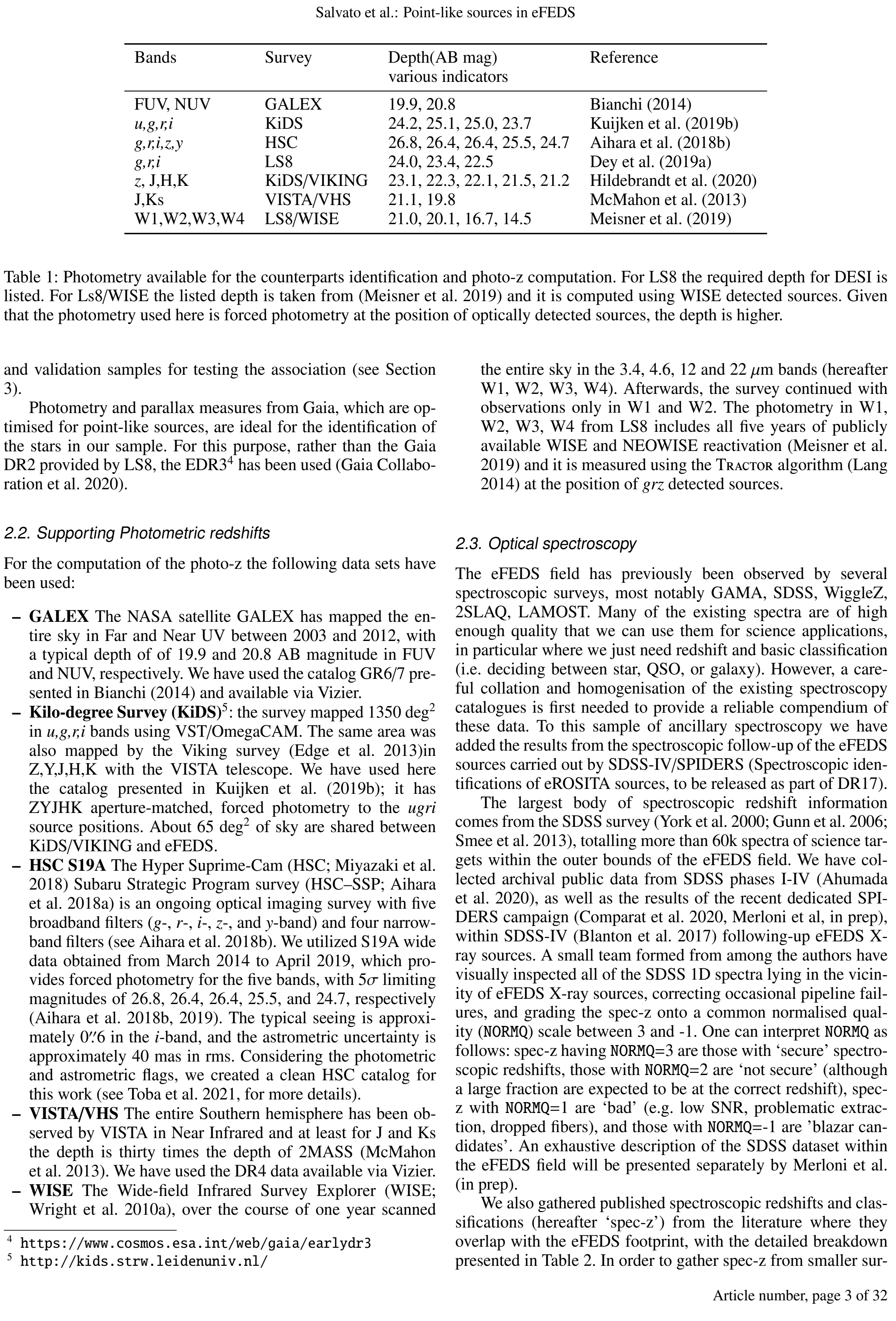
The identification of the counterparts is carried out using the DESI Legacy Imaging Survey DR8 (LS8; Dey et al. 2019b), for various reasons. First of all, it covers homogeneously the field and it has sufficient depth, based on the expectation of the X-ray population optical properties (Merloni et al. 2012). In addition, the survey provides data that are registered to Gaia DR2 (i.e. higher positional precision) and provides, together with Gaia, also the AllWISE *tractor* (Lang 2014) photometry extracted at the position of the optical sources. Finally, the survey covers 14,000 square degrees of sky thus providing a sufficient number of sources external to eFEDS that can be used as training

---

[1] this parameter is obtained from the task srctool of the eSASS software; Brunner et al., 2021, submitted

[2] i.e., comparable to the XMM Slew Survey: `https://www.cosmos.esa.int/web/xmm-newton/xmmsl2-ug`

[3] `http://dr5.lamost.org/doc/release-note-v3`



Bands	Survey	Depth(AB mag) various indicators	Reference
FUV, NUV	GALEX	19.9, 20.8	Bianchi (2014)
u,g,r,i	KiDS	24.2, 25.1, 25.0, 23.7	Kuijken et al. (2019b)
g,r,i,z,y	HSC	26.8, 26.4, 26.4, 25.5, 24.7	Aihara et al. (2018b)
g,r,i	LS8	24.0, 23.4, 22.5	Dey et al. (2019a)
z, J,H,K	KiDS/VIKING	23.1, 22.3, 22.1, 21.5, 21.2	Hildebrandt et al. (2020)
J,Ks	VISTA/VHS	21.1, 19.8	McMahon et al. (2013)
W1,W2,W3,W4	LS8/WISE	21.0, 20.1, 16.7, 14.5	Meisner et al. (2019)

Table 1: Photometry available for the counterparts identification and photo-z computation. For LS8 the required depth for DESI is listed. For Ls8/WISE the listed depth is taken from (Meisner et al. 2019) and it is computed using WISE detected sources. Given that the photometry used here is forced photometry at the position of optically detected sources, the depth is higher.

and validation samples for testing the association (see Section 3).

Photometry and parallax measures from Gaia, which are optimised for point-like sources, are ideal for the identification of the stars in our sample. For this purpose, rather than the Gaia DR2 provided by LS8, the EDR3[4] has been used (Gaia Collaboration et al. 2020).

2.2. Supporting Photometric redshifts

For the computation of the photo-z the following data sets have been used:

- **GALEX** The NASA satellite GALEX has mapped the entire sky in Far and Near UV between 2003 and 2012, with a typical depth of of 19.9 and 20.8 AB magnitude in FUV and NUV, respectively. We have used the catalog GR6/7 presented in Bianchi (2014) and available via Vizier.
- **Kilo-degree Survey (KiDS)**[5]: the survey mapped 1350 deg^2 in *u,g,r,i* bands using VST/OmegaCAM. The same area was also mapped by the Viking survey (Edge et al. 2013)in Z,Y,J,H,K with the VISTA telescope. We have used here the catalog presented in Kuijken et al. (2019b); it has ZYJHK aperture-matched, forced photometry to the *ugri* source positions. About 65 deg^2 of sky are shared between KiDS/VIKING and eFEDS.
- **HSC S19A** The Hyper Suprime-Cam (HSC; Miyazaki et al. 2018) Subaru Strategic Program survey (HSC–SSP; Aihara et al. 2018a) is an ongoing optical imaging survey with five broadband filters (*g*-, *r*-, *i*-, *z*-, and *y*-band) and four narrow-band filters (see Aihara et al. 2018b). We utilized S19A wide data obtained from March 2014 to April 2019, which provides forced photometry for the five bands, with 5σ limiting magnitudes of 26.8, 26.4, 26.4, 25.5, and 24.7, respectively (Aihara et al. 2018b, 2019). The typical seeing is approximately $0\farcs6$ in the *i*-band, and the astrometric uncertainty is approximately 40 mas in rms. Considering the photometric and astrometric flags, we created a clean HSC catalog for this work (see Toba et al. 2021, for more details).
- **VISTA/VHS** The entire Southern hemisphere has been observed by VISTA in Near Infrared and at least for J and Ks the depth is thirty times the depth of 2MASS (McMahon et al. 2013). We have used the DR4 data available via Vizier.
- **WISE** The Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010a), over the course of one year scanned the entire sky in the 3.4, 4.6, 12 and 22 μm bands (hereafter W1, W2, W3, W4). Afterwards, the survey continued with observations only in W1 and W2. The photometry in W1, W2, W3, W4 from LS8 includes all five years of publicly available WISE and NEOWISE reactivation (Meisner et al. 2019) and it is measured using the Tractor algorithm (Lang 2014) at the position of *grz* detected sources.

2.3. Optical spectroscopy

The eFEDS field has previously been observed by several spectroscopic surveys, most notably GAMA, SDSS, WiggleZ, 2SLAQ, LAMOST. Many of the existing spectra are of high enough quality that we can use them for science applications, in particular where we just need redshift and basic classification (i.e. deciding between star, QSO, or galaxy). However, a careful collation and homogenisation of the existing spectroscopy catalogues is first needed to provide a reliable compendium of these data. To this sample of ancillary spectroscopy we have added the results from the spectroscopic follow-up of the eFEDS sources carried out by SDSS-IV/SPIDERS (Spectroscopic identifications of eROSITA sources, to be released as part of DR17).

The largest body of spectroscopic redshift information comes from the SDSS survey (York et al. 2000; Gunn et al. 2006; Smee et al. 2013), totalling more than 60k spectra of science targets within the outer bounds of the eFEDS field. We have collected archival public data from SDSS phases I-IV (Ahumada et al. 2020), as well as the results of the recent dedicated SPIDERS campaign (Comparat et al. 2020, Merloni et al, in prep), within SDSS-IV (Blanton et al. 2017) following-up eFEDS X-ray sources. A small team formed from among the authors have visually inspected all of the SDSS 1D spectra lying in the vicinity of eFEDS X-ray sources, correcting occasional pipeline failures, and grading the spec-z onto a common normalised quality (`NORMQ`) scale between 3 and -1. One can interpret `NORMQ` as follows: spec-z having `NORMQ`=3 are those with 'secure' spectroscopic redshifts, those with `NORMQ`=2 are 'not secure' (although a large fraction are expected to be at the correct redshift), spec-z with `NORMQ`=1 are 'bad' (e.g. low SNR, problematic extraction, dropped fibers), and those with `NORMQ`=-1 are 'blazar candidates'. An exhaustive description of the SDSS dataset within the eFEDS field will be presented separately by Merloni et al. (in prep).

We also gathered published spectroscopic redshifts and classifications (hereafter 'spec-z') from the literature where they overlap with the eFEDS footprint, with the detailed breakdown presented in Table 2. In order to gather spec-z from smaller sur-

---

[4] `https://www.cosmos.esa.int/web/gaia/earlydr3`

[5] `http://kids.strw.leidenuniv.nl/`

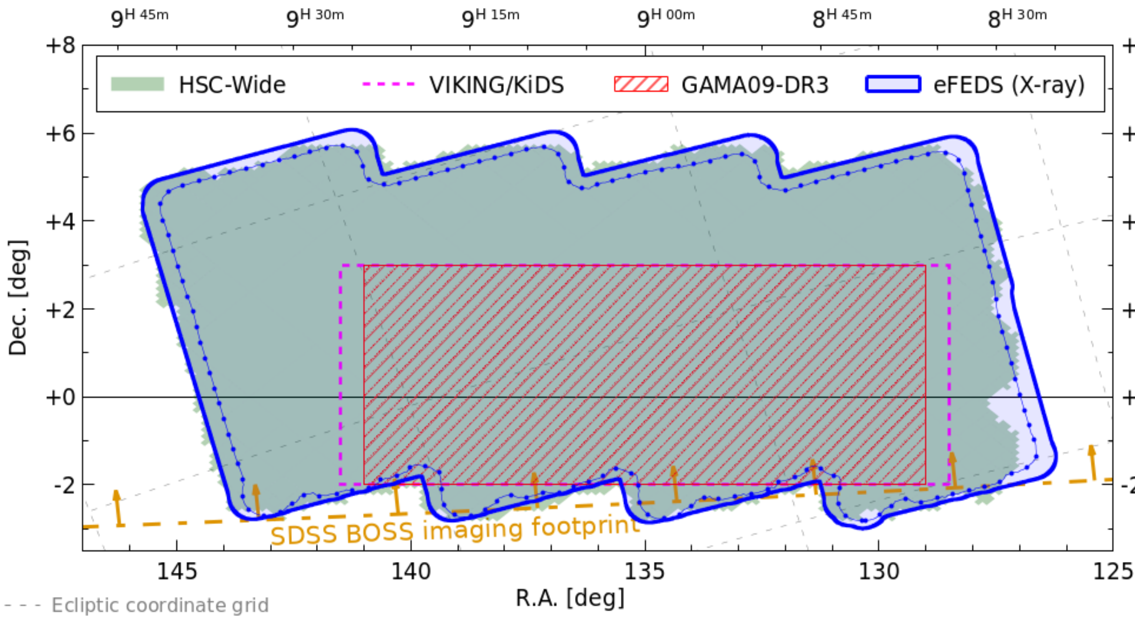


Fig. 1: eFEDS X-ray and multi-wavelength coverage. The thick blue line shows the outer bound of the region that was searched for X-ray sources. The thin blue beaded line shows the region having at least 500 seconds of effective X-ray exposure depth. We indicate the approximate coverage of several selected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

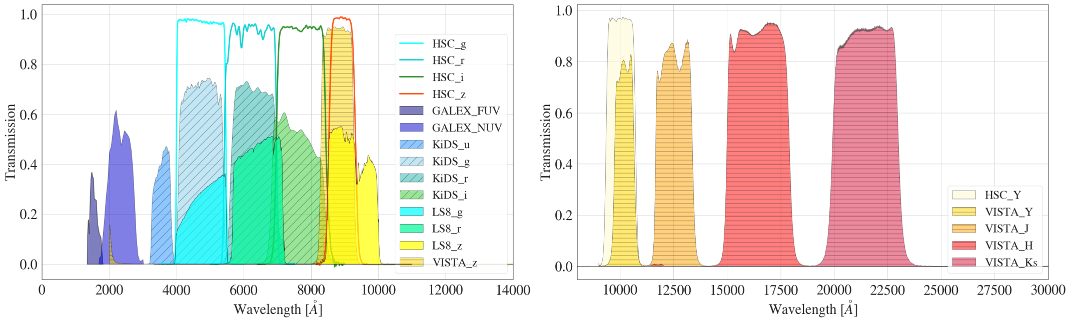


Fig. 2: Illustration of the band-pass and relative transmission curves of the UV/optical/NIR photometry used for computation of photo-z in this work. For clarity, we do not show the WISE band-passes.

veys that might only contribute a few redshifts each, we have also queried the Simbad database (as of 05/03/2021, Wenger et al. 2000), in the vicinity of the eFEDS X-ray positions.

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

All these spec-z were collated into a single catalogue, with a single redshift and classification per sky position, using a match in coordinates between 1-3 arcsec , depending on the input target catalogue. After the de-duplication step we are left with 143637 unique entries over the eFEDS field, of which 108834 are 'secure' (i.e `NORMQ`=3).

3. Counterparts identification: Methodology

Because of the large PSF of the eROSITA telescopes and the small number of photons associated with typical X-ray detections, the 1σ rms positional uncertainties of individual X-ray sources can be several arcseconds. Specifically, in eFEDS the

Spectroscopic Survey	Quality threshold in original catalog	Number of Sources	Data Release & Reference
SDSS	non-null redshift	60k	(up to DR17; Merloni et al., in prep)
GAMA	NQ>=4	26318	(DR3, Baldry et al. 2018)
WiggleZ	Q>= 4	13466	(Final DR, Drinkwater et al. 2018b)
2SLAQ	q_z2S=1	953	(v1.2, Croom et al. 2009)
6dFGS	4 ≤ q_cz ≤	365	(final data release, Jones et al. 2009)
2MRS	non-null redshift	152	(v2.4, Huchra et al. 2012)
LAMOST	snrr> 10, z>= −1.0, 0.0 < z_err < 0.002	55866	(DR5 v3, Luo et al. 2015)
Gaia RVS	non-null redshift	15568	(DR2, Gaia Collaboration et al. 2018)
Simbad	non-null redshift	3915	(as of 05/03/2021, Wenger et al. 2000)

Table 2: Spectroscopic redshifts available within the eFEDS footprint (126<RA<146.2 deg, -3.2<Dec<+6.2 deg). For Simbad, the number of entries is limited to objects lying within 3 arcsec of the optical coordinates of counterparts to eFEDS sources.

mean positional error is 4″.7 and extends above 20 arcsec only for a handful of sources; see Brunner et al., (submitted). For the expected optical/infrared magnitude distribution of X-ray sources at the depth of the eFEDS (see e.g., Merloni et al. 2012; Menzel et al. 2016), the sky density of the relevant astrophysical source populations is often large. For this reason, the identification of the true associations cannot be determined solely by closest neighbour searches, as there will be several potential counterparts within the error circle of any given X-ray source. Taking this into account, the identification of the counterparts of eFEDS point-like sources has been performed using two independent methods.

NWAY (Salvato et al. 2018b), based on Bayesian statistics, and ASTROMATCH (Ruiz et al. 2018) based on the Maximum Likelihood Ratio (MLR; Sutherland & Saunders 1992), have been specifically developed to identify the correct counterparts to X-ray sources, independently of their Galactic or extragalactic nature. In order to assess the probability (or likelihood) of an object to be the correct counterpart to an eFEDS sources, the two methods take first into account the separation between the sources, their positional accuracy and the number density of the sources in the ancillary data. The difference between the methods resides then in the adoption of different assumptions (priors) on the properties of the counterparts and thus on the multi-wavelength training samples used. After the identification of the key features, we have applied the two algorithms to a blind validation sample.

In the following, we describe first the construction of the training, validation and associated field samples. These are the samples that NWAY and ASTROMATCH have used for identifying the several but mostly different features considered to be relevant for the determination of the counterparts. The description of the two methods ends the section.

3.1. Construction of the training, validation and field samples

To judge between two counterparts at equal distance from an X-ray source, one may consider how typical for example the magnitude is for an X-ray emitter, compared to how typical it is for an unrelated source. This idea of discriminating based on magnitude distributions was extended in Salvato+18 in NWAY to a two-dimensional color-magnitude distribution. Here, we go further and use machine learning to discriminate target and field distributions based on many bands. To train a machine learning methods, a clean sample of the X-ray source population representative for the depth and galactic latitude of eFEDS is needed. Expanding on the methods used to determine counterparts to the ROSAT all-sky survey (Dwelly et al. 2017; Salvato et al. 2018b), we have exploited serendipitous X-ray catalogues (3XMM, CSC2; Rosen et al. 2016; Evans et al. 2010), together with LS8 (Dey et al. 2019a), in order to construct clean reference samples of point-like X-ray sources that have secure optical/IR (OIR) counterparts and that span the X-ray flux range probed by eFEDS. These reference samples from 3XMM and CSC2 are used to train and validate respectively, the cross-matching algorithms used to correctly pair the eFEDS X-ray sample to OIR counterparts. Our primary goal when creating these reference samples is to include only a very clean subset i.e. only those that have sufficiently well constrained X-ray positions such that their association with a single OIR counterpart is trivial, regardless of the Galactic or extragalactic nature. The parent X-ray catalogues are large enough (few 10^5 objects) that we can apply stringent quality filters but still obtain a sufficiently large pool of X-ray/OIR reference sources. We also construct representative 'field' samples from the OIR catalogue, these are used later to inform our cross-matching algorithms about the (majority of) stars and galaxies that are *not* emitting X-rays at the fluxes probed by the eFEDS survey.

3.1.1. Reference sample selected from the 3XMM-DR8 serendipitous source catalogue (a.k.a. training sample)

We start with the 3XMM-DR8[6] catalogue of X-ray detections, and estimate the 0.5-2 keV X-ray flux (and uncertainty) of each detection from the 0.5-1, and 1-2 keV band fluxes (and their uncertainties). We then select only those detections that meet all of the following X-ray quality criteria:

i. have X-ray flux in the range probed by eFEDS ($F_{0.5-2\mathrm{keV}} > 2 \times 10^{-15}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$),
ii. have detection likelihood > 10,
iii. have X-ray positions that have been aligned with the optical frame and that have uncertainty smaller than 1.5 arcsec,
iv. have a signal to noise ratio for $F_{0.5-2\mathrm{keV}}$ that is greater than 10,
v. are consistent with being point-like at the resolution of *XMM-Newton*,
vi. have no close X-ray neighbours within 10 arcsec,
vii. were not detected at the extreme off-axis angles,
viii. were detected in *XMM-Newton* exposures of at least 5 ks and finally,

---

[6] `http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR8/3XMM_DR8.html`

ix. were not labelled by the 3XMM pipeline as being confused, affected by high X-ray background or flagged as being problematic for any reason.

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

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

After applying these criteria, we are left with a sample of 36276 unique point-like X-ray sources with median positional uncertainty 0.57 arcsec . The X-ray flux distribution of the 3XMM-DR8-based training sample is broadly similar to that of the science sample; 92% of the sample have 0.5-2keV fluxes in the range 5x10^{-15} - 1x10^{-12} cgs, and median flux is 1.7x10^{-14} cgs.

We then carried out a positional match of this X-ray sample to the LS8, considering optical/IR objects that lie within 5 arcsec of the X-ray positions. We used NWAY (Salvato et al. 2018b) to carry out this cross-match, using only astrometric information and number densities (NWAY basic mode, i.e. without any magnitude or colour priors). We retain only the X-ray sources with very secure unique optical counterparts. Specifically, we require that we consider only X-ray sources with >90% probability of having an optical/IR counterpart, and only cases where the best optical/IR counterpart is at least 9 times more probable than the next best possibility (`p_any` > 0.9, `p_i` > 0.9)[7]. As before, we can afford to be very strict with these criteria, since we primarily care about purity and not completeness. These cuts result in a 3XMM/LS8 reference sample of 20705 high quality X-ray/OIR matches.

We select a corresponding sample of non-X-ray emitting field objects from the LS8, using annular regions (15, 30 arcsec radii) around each of the 20 705 reference sample positions. The field sample was further filtered to remove any object that lies within 15 arcsec of *any* 3XMM-DR8 source. This field sample contains just under 396 000 entries.

3.1.2. Reference sample selected from the Chandra Source Catalogue v2.0 (a.k.a. validation sample)

A supplementary X-ray/OIR reference sample was derived from the Chandra Source Catalogue v2.0[8]. We used the Web API to retrieve all CSC2 sources that satisfied the following X-ray quality criteria, regardeless their position in the sky: i) Have $F_{0.5-2\mathrm{keV}} > 2\times10^{-15}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ (estimated from the standard CSC2 ‘s’ and ’m’ bands), ii) have high significance > 6, iii) have a signal to noise ratio on $F_{0.5-2\mathrm{keV}}$ that is greater than 5, iv) have X-ray positions with 95% uncertainty ellipse radius smaller than 1.0 arcsec , v) are consistent with being point-like at the resolution of Chandra, vi) were detected in Chandra exposures of at least 1 ks and finally, vii) were not labelled by the CSC pipeline as being confused, affected by readout streaks, or piled up. Exactly the same sky region filtering criteria were applied to the CSC-based sample as were used to filter the 3XMM-based reference sample (see section 3.1.1). These criteria result in a sample of 6066 clean reference X-ray sources.

We followed a similar process as before (section 3.1.1) to match the CSC2 sources to the LS8 catalogue, retaining only unassailable matches (having `p_any` > 0.9, `p_i` > 0.9). This results in a CSC-based X-ray/OIR reference sample that contains 3415 objects.

As before, we selected a corresponding LS8 field sample, from annular regions around each reference source, excluding any entries that lie within 15 arcsec of any CSC2 source having $F_{0.5-2\mathrm{keV}} > 5\times10^{-16}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$. This field sample contains just over 67 000 objects.

3.2. A machine-learning based approach to photometric priors for NWAY

In addition to astrometry, i.e. the separation between an X-ray source and a candidate counterpart, the associated positional uncertainties and the number densities of the sources in the two catalogs, the photometry of potential counterparts is valuable information to determine whether or not they are associated to a given X-ray detection. Traditionally, the likelihood ratio associated to angular distance was multiplied by a factor accounting for the magnitude distributions and the sky density of a population of X-ray sources and background objects (e.g Brusa et al. 2005, 2007; Luo et al. 2010). In NWAY, this idea was re-formulated in the Bayesian formalism, in the following way.

Given some data D, the posterior association probability $P(H \mid D)$ is related to the prior probability of chance alignment $P(H)$ via the likelihood $P(D \mid H)$, $P(H \mid D) \propto P(H) \times P(D \mid H)$. If photometric information (or any other feature, in fact) is used, then the likelihood becomes: $P(D \mid H) = P(D_\phi \mid H) \times P(D_m \mid H)$ where D_ϕ and D_m refer to the astrometric and photometric information, respectively. For any possible association, the modifying factor $P(D_m \mid H)$ is computed from the feature (e.g., magnitude or colour) m of the counterpart candidate and from the expected distribution of this observable for X-ray sources and field (non X-ray) sources. We call such factors "priors" to NWAY, as they enter as *a priori* information in the ultimate matching process. These priors are posteriors previously learned from other data. For further details on the formalism we refer to Salvato et al. (2018b) and the NWAY documentation[9].

In order to take full advantage of the LS8 ancillary catalogue, we have extended this approach for the eFEDS counterpart identification. Instead of using a subset of magnitudes, colours and their associated distributions, we have trained a Random Forest classifier (`sklearn` implementation, Pedregosa et al. 2011) on a large number of features to reliably map the available Legacy DR8 information to real X-ray sources and real field objects. The trained classifier is then used to predict the probability of all counterpart candidates to be X-ray emitting. This probability is directly used to compute $P(D_m \mid H)$. In the following section, we describe the definition of the features in the training sample.

---

[7] in NWAY `p_any` is the probability, for each source in the primary catalog (eFEDS in this case) to have a counterpart in the secondary catalogs; then, for each source in the secondary catalogues, `p_i` gives the probability to be the correct counterpart to the source in the primary catalogue (see more in the NWAY manual and Salvato et al. (2018b)

[8] `https://cxc.cfa.harvard.edu/csc2/index.html`

[9] `https://github.com/JohannesBuchner/NWAY`

3.2.1. Random Forest prior: training and performance

From the 3XMM training sample described above, we have extracted a set of photometric and astrometric features. The training features are listed and described in Table 3. X-ray sources are flagged as target class "1", field objects as target class "0". 15% of the randomized samples are extracted for testing purposes and not further considered in the training procedure. The baseline model is composed of 200 trees, allowing decision split points if at least 8 samples are left in each branch. All of the 22 features can be used for the decision-tree building, which makes use of bootstrap samples of the training set.

By construction the training sample is highly imbalanced, since the field objects strongly outnumber the X-ray sources. We therefore opted for a weighting scheme, automatically adjusting weights of training examples for the class imbalance. The trained model is evaluated on the test set, resulting in the confusion matrix presented in Figure 3. We note that the cut in the class prediction for the presented confusion matrix is made at $p_{\rm X-ray} = 0.50$, where $p_{\rm X-ray}$ is the predicted probability that a counterpart candidate is X-ray emitting. Since NWAY uses the continuous predicted probability as modifying factor for the likelihood $P(D \mid H)$, real counterparts with rare or untypical photometric features, i.e with $p_{\rm X-ray} \lesssim 0.50$, may still be selected by the algorithm if the astrometric configuration favours them. We obtain a high recall fraction of $2517/(2517 + 471) = 0.84$, while the fractional leakage of contaminating field objects remains low: $782/(782 + 58051) = 0.01$.

Feature	Description
flux_*/mw_transmission_*	deredenned flux in *g,r,z,W1,W2*
gaia_phot_*_mean_mag	original GAIA phot. in G, $G_{\rm bp}$, $G_{\rm rp}$
snr_*	S/N for *g,r,z,W1,W2,G*,$G_{\rm bp}$,$G_{\rm rp}$
$\sqrt{pmra^2 + pmdec^2}$	Gaia proper motion
parallax	Gaia paralllax
g-r, r-z, z-W1, r-W2	dereddened colors

Table 3: LS8 training features used to model the photometric prior.

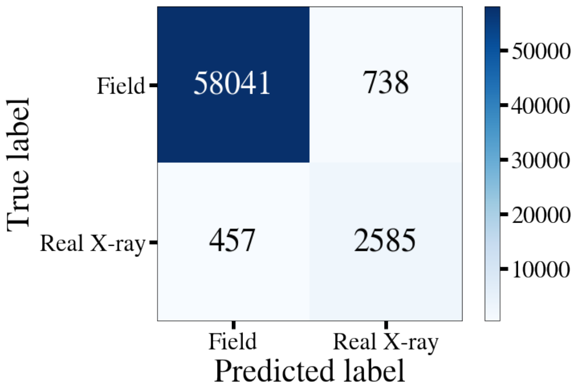


Fig. 3: Confusion matrix resulting from the random forest prediction on an independent test set. X-ray sources are labelled as "Real X-ray" while field objects as "Field". Numbers on the right downward diagonal correspond to correctly predicted classes.

3.2.2. NWAY association run

Using the trained model, we predict $p_{\rm X-ray}$ for all LS8 sources in the eFEDS field. We then run the NWAY matching procedure using $p_{\rm X-ray}$ for the modifying factor to the likelihood function. This is done by adding $p_{\rm X-ray}$ as a column to the LS8 catalog and activating it as a prior column in NWAY. We set a radius of 30 arcsec from each eFEDS X-ray source, considering all LS8 sources within this radius. This relatively large maximal separation accounts for the largest positional uncertainties of a few objects in the eFEDS source catalogue and the use of a large search radius minimises the probability of missing counterparts that are widely separated from the X-ray centroid position. The sky coverage of eFEDS and LS8 are respectively 140 deg^2 and $N_{\rm eFEDS} \times \pi \times (30'')^2 - A_{\rm overlap}$ where $N_{\rm eFEDS}$ is the number of eFEDS sources (point-like or extended) and $A_{\rm overlap}$ the overlap area of neighbouring search windows around the X-ray sources.

3.3. The MLR approach

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

$$LR = \frac{q(\overrightarrow{m}) \cdot f(\mathbf{r})}{n(\overrightarrow{m})}, \tag{1}$$

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

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

The version of MLR applied to the eFEDS work is based on the ASTROMATCH[10] implementation. This tool has been specifically designed to deal with the complexity of wide-area surveys that contain a very large number of sources. The HEALPix Multi-Order Coverage map (MOC[11]) technology is used to describe the footprint of a catalogue of astrophysical sources. The KD-tree library as implemented in the ASTROPY package (Astropy Collaboration et al. 2013, 2018) is used to accelerate spatial searches of potential counterparts within a radius **r** of a

---

[10] `https://github.com/ruizca/astromatch`

[11] `https://www.ivoa.net/documents/MOC/`

given sky position. The core ASTROMATCH functionality is expanded to enable the use of multidimensional priors. The version of ASTROMATCH adopted in this work is therefore a fork (`github.com/ageorgakakis/astromatch`) of the main development branch.

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

The priors are generated using the 3XMM training sample of Subsection 3.1. The LS8 photometric properties of the sources in that sample were explored to identify parameter spaces, in which they separate from the general LS8 field population. After some experimentation we opted for the following 3 independent priors:

- A space that includes the WISE colour $W1 - W2$, the WISE magnitude $W2$ and the optical extent of a source. For the latter we use the LS8 parameter TYPE, which provides information on the optical morphology of sources. In our application we only differentiate between optically unresolved (TYPE="PSF") and optically extended (TYPE≠"PSF") populations.
- A space that includes the optical/WISE colour $r - W2$, the optical magnitude g and the optical extent of a source. For the latter we use the Legacy-DR8 parameter TYPE as explained above.
- The distribution of the Gaia G magnitudes listed in the LS8 catalogues. This is to identify X-ray sources associated with very bright counterparts.

The distribution of the training sample sources in the parameter spaces above is used to define two 3-dimensional and one 1-dimensional independent priors. These are provided as input to the ASTROMATCH code, together with the distribution of the sources in the field population. For a given X-ray source all the potential associations within the search radius of 30 arcsec are identified. Each of them is assigned one LR value for each of the 3 priors using Equation 1. The highest of the three LR values is assigned to the source.

4. Results of identification methods, comparison and merging

NWAY and ASTROMATCH are run independently on the 27369 sources in the eFEDS point-source catalogue. The results are then consolidated and a counterpart reliability flag (CTP_quality) is assigned to each source. The same process is then repeated for the 246 sources in the eFEDS hard point-source catalog. From now on, all numbers and descriptions are given for the main sample, unless specified otherwise.

After the consolidation of the counterpart, a further test for consistency is done by comparing the results of the association with an independent method, HamStar (Schneider et al., submitted), which is tuned on the identification of Galactic coronal X-ray emitters (Section 4.5). The details on the steps are provided below.

4.1. The eROSITA-like validation sample

The validation sample of 3415 counterparts to *Chandra* sources (see Section 3.1) was used as a truth table to test the performance of NWAY and ASTROMATCH for finding counterparts and to define the `p_any` and `LR_BEST` thresholds for the association.

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

Instead of directly using the assigned σ as the astrometric error to be applied to the Chandra positions to make them resemble the eFEDS astrometric accuracy, we prefer to add further randomness to the experiment. For each *Chandra* source the assigned $\sigma/\sqrt{2})$ is treated as the scale factor of the Rayleigh distribution and a deviate is drawn, which represents the positional error. This is applied to the sky coordinates of the optical counterpart of the *Chandra* source and the new offset position is taken as the centroid of the X-ray source in the case of an eFEDS-like observation.

4.2. Probability thresholds definition

The X-ray positions of the *Chandra* eFEDS-like sources in the validation sample are matched to the LS8 multi-wavelength dataset using the same setup for the NWAY and ASTROMATCH methods used with the real eFEDS observations. The resulting catalogue of best counterparts can be compared with the true association of each of the 3415 *Chandra* sources. This approach provides a measure of the false-positive identification rate of the eFEDS counterpart catalogue. We compared the primary identifications returned by NWAY and ASTROMATCH to true identifications stored in the validation sample. At any given value of `p_any`/`LR_BEST` we define as *purity* the fraction of sources with the correct identification. In addition we define as *completeness* the fraction of sources for which we can assign a counterpart, The curves of purity and completeness as a function of `p_any` and `LR_BEST` are shown in Fig. 4. The thresholds are defined at the point in which completeness and purity cross each other (see e.g., Marchesi et al. 2016). This corresponds to 0.035 for `p_any` and 0.45 for `LR_BEST`. NWAY and ASTROMATCH return 23980 and 23415 primary counterparts above threshold, respectively.

4.3. Comparison of counterparts from NWAY and ASTROMATCH

For 24193/27369 (88.4%) eFEDS point like sources in the main sample, NWAY and ASTROMATCH point at the same counterpart and disagree for 3176 (11.6%) of the cases.

Table 4 summarises the number of eFEDS sources with the agreement/disagreement between the two methods as a function of detection likelihood of the X-ray source. Sources with low detection likelihood values have, on average, larger X–ray positional errors and a larger number of spurious sources is expected from simulations (Brunner et al., submitted, Liu et al. submitted). It is therefore not surprising that the largest discrep-

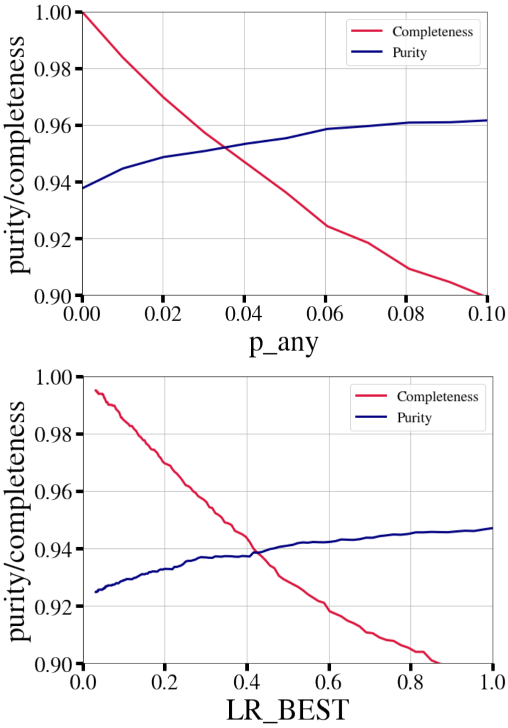


Fig. 4: Purity (blue solid line) vs. completeness (red solid line) as a function of `p_any` for the association of the CSC2 eROSITA-like validation sample made with NWAY (top panel) and as a function of `LR_BEST`, for the ASTROMATCH (bottom panel).

Sample	Number	Counterparts	
		"same"	"different"
DET_LIKE_0> 6	27369	24193	3176 (11.6%)
DET_LIKE_0> 8	21410	19162	1795 (8.4%)
DET_LIKE_0> 10	17574	16435	1136 (6.5%)

Table 4: Comparison of matches between NWAY and ASTROMATCH. In the last column the fraction of the "different ctps" with respect to the whole sample is also reported.

ancies are observed at the lowest detection likelihoods (Fig. 5). In fact, the disagreement drops from 11.6% to 6.5% when considering only eFEDS sources with `DET_LIKE` grater than 10, suggesting that at low detection likelihood a fraction of eFEDS sources might be spurious detections where NWAY and ASTROMATCH assign a different "field" source. The notion that these are "field" sources is also supported by the fact that for about 50% of eFEDS sources with `DET_LIKE` below 10 and with different counterparts, both `p_any` and `LR_BEST` are below threshold.

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

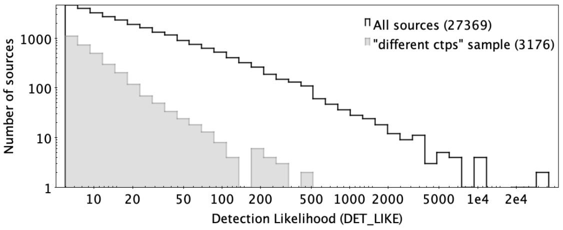


Fig. 5: Number of sources as a function of detection likelihood for the entire sample of eFEDS Main catalog sources (black histogram) and for the sources with NWAY and ASTROMATCH indicating different counterparts (grey shaded histogram).

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

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

4.4. Assigning a quality to the proposed counterpart

As a consequence of the discussion above, each counterpart in the catalog has been flagged as following ([number] refers to the number of sources in the category):

- `CTP_quality`=4: when NWAY and ASTROMATCH agree on the counterpart and both `p_any` and `LR_BEST` are above threshold [20873 sources];
- `CTP_quality`=3: when NWAY and ASTROMATCH agree on the counterpart but only one of the methods assigns the counterpart with a probability above threshold [1379];
- `CTP_quality`=2: when there is more than one possible reliable counterpart. This includes a) all the sources in the "different ctps" sample with at least one probability above threshold and b) the sources in the "same ctps" sample with possible secondary solutions [2522 in total]. This latter case, due to the low spatial resolution of eROSITA, implies that both sources are contributing to the X-ray flux. A supplementary catalog with the properties of the secondary counterparts for these 2522 sources is also released (Section 7).
- `CTP_quality`=1: when NWAY and ASTROMATCH agree on the counterpart but both `p_any` and `LR_BEST` are below threshold [1370]; Note that a probability below threshold does not necessarily imply a wrong counterpart. It could also indicate that the counterpart is correct but its features are not sufficiently mimicking those in the training sample.

"Same ctps", single solutions – 23622 (86.3%)		
	`p_any` >0.035	`p_any` <0.035
`LR_best`> 0.45	**20873 (0.763)**	**561 (0.020)**
`LR_best`< 0.45	**818 (0.030)**	**1370 (0.050)**
"Same ctps", multiple solutions – 571 (2.1%)		
	`p_any` >0.035	`p_any` <0.035
`LR_best`> 0.45	**505 (0.018)**	**7 (3x10^{-4})**
`LR_best`< 0.45	**59 (0.002)**	-
"Different ctps" – 3176 (11.6%)		
	`p_any` >0.035	`p_any` <0.035
`LR_best`> 0.45	**1243 (0.045)**	**226 (0.008)**
`LR_best`< 0.45	**478 (0.017)**	**1225 (0.045)**
HamStar[1]	**4 (2x10^{-4})**	-

Table 5: Counterparts quality summary. Comparison of matches with NWAY and ASTROMATCH as a function of their respective thresholds, split between "same counterparts" (for both cases of single and multiple counterparts) and "different counterparts" classes; in parenthesis the fractions of the total sample. The numbers in each box are color-coded by their `CTP_quality` value (see text for more details): Thick black = 4; thick blue = 3; thick cyan = 2; thick purple = 1 and thick red = 0. [1] Objects for which HamStar would point to a different counterpart with `p_stellar`>0.95 are by definition sources with `CTP_quality`=2 (see text for details).

- `CTP_quality`=0: when NWAY and ASTROMATCH indicate different counterparts and both `p_any` and `LR_BEST` are below threshold [1225].

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

4.5. Validation with an independent association method tuned to stars: HamStar

The content of the eFEDS point-source catalog has also been analyzed in order to specifically identify stellar coronal X-ray emitters with sufficiently well-defined properties. This method, called 'HamStar' in the following, is based on the properties expected for this type of star; the details are presented in Schneider et al. (submitted). In short, HamStar performs a binary classification between stellar coronal emitters and other objects. This classification is based on the concept of eligible stellar counterparts, i.e., the match catalog contains only stellar objects that may reasonably be responsible for the X-ray sources. Specifically, the parent sample that HamStar uses includes only sources from Gaia EDR3 that:

- are brighter than 19th magnitude in G band (implied by the stellar saturation limit of $L_X/L_{\rm bol} \lesssim 10^{-3}$ and the depth of eFEDS);
- have accurate magnitudes in all three Gaia photometric bands (to apply color-dependent corrections);
- have a parallax value at least three times larger than the parallax error (to select only genuine stars).

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

Of the 2060 eFEDS sources with a counterpart from HamStar, 1842 have the counterpart identified in this work that is less than 2 arcsecfrom the counterpart proposed by Hamstar and we assume to be the same source. We visually inspected the cutouts of the 29 sources for which the separation between the counterpart proposed by Hamstar and this work is between 2 and 3 arcsec, and concluded that for 9 sources the counterparts are the same but the sources are heavily saturated in LS8 so that the coordinates are not sufficiently precise. This corresponds to an 89.8% agreement; incidentally, this value corresponds almost exactly to the expected reliability and completeness of HamStar (Schneider et al., submitted). All these sources will be then classified as "Secure galactic" in Section 5.

HamStar applies well-understood X-ray-to-optical properties of stars to a well-defined subsample of Gaia sources. On the other hand, the training samples used by NWAY and ASTROMATCH include various classes of X-ray emitters: stars and compact objects, AGN and galaxies, including the bright ones at the center of clusters (BCG). Thus, we decided to keep the counterparts proposed by NWAY/ASTROMATCH also for the remaining 211 cases, for which HamStar would point to a different source and change the quality flag if needed; for 106/211 sources already the `CTP_quality` was set to 2. For four additional sources for which HamStar has a probability to point to the right counterpart higher than 95%, we degrade the `CTP_quality` flag from 3 to 2. For the 101 remaining sources `CTP_quality` was set to 0 or 1 and such it remains.

The eFEDS field spans a wide range of Galactic latitudes (from about +20 to about +40). Reassuringly, the fraction of the X-ray sources which are classified as Galactic (see Section 3 for the definition) increases towards low Galactic latitude, as shown in Figure 6. The figure also confirms that the priors adopted by Hamstar are not representative of all Galactic sources in eFEDS. In fact about 22.3% of the Galactic sources identified by NWAY/ASTROMATCH are fainter than the 19th magnitude (with 10% Gaia non detected).

4.6. Separation and magnitude distribution of the counterparts

For 24427/24774 (98.5%) of the sources having `CTP_quality`≥2, the separation between the X-ray position and the assigned LS8 counterpart is smaller than 15 arcsec, with a mean of 4.3 arcsec. As might be expected, there is a trend for larger average X-ray-optical separations at smaller values of DET_LIKE; lower detection likelihood sources typically have larger X-ray positional uncertainty (see Brunner et al., 2021, submitted). The distribution of the observed X-OIR separations normalised by the X-ray positional uncertainty is shown in Figure 7 as function of the r magnitude of the counterpart.

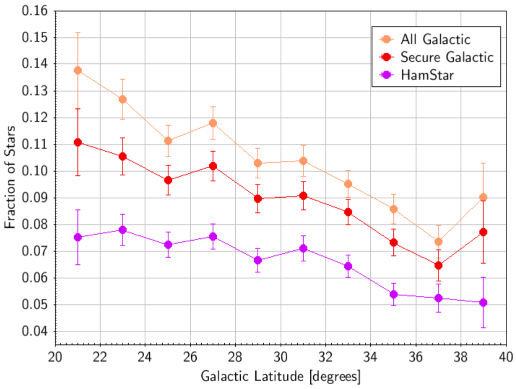


Fig. 6: Fraction of all X-ray sources with `CTP_quality`≥2 which are classified as Galactic as a function of Galactic Latitude. Purple symbols are for objects classified by the HamStar method; red symbols are for all "Secure Galactic" objects and orange for both "Secure" and "Likely" Galactic objects (see Section 3 for the definition).

The distribution is broadly comparable to the expectation of a Rayleigh distribution with scale factor = 1.

In Figure 8 we show the distribution of the sample in X-ray flux versus optical magnitude space, with the sample subdivided into those objects with more secure counterparts (`CTP_quality`≥2), and those with less reliable counterparts (`CTP_quality`≤1). The less reliable counterparts tend to have fainter optical magnitudes for a given X-ray flux than the more secure counterparts.

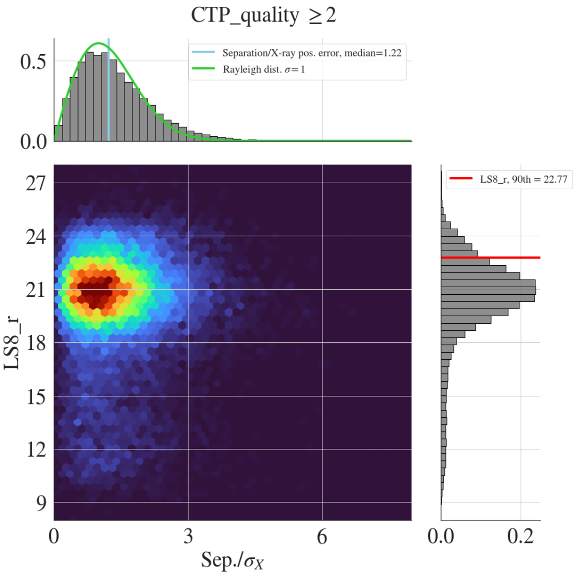


Fig. 7: Separation between X-ray position and the selected counterpart normalized by the 1-dimensional positional error of the X-ray source, as a function of its r band magnitude for the sources with secure counterparts (`CTP_quality`≥2). The hexagons are color coded linearly according to the counts in the specific bins. The marginal histograms have linear y-axis. The 90-th percentile of the r-band magnitude distribution (22.77) and the median of the normalised X-OIR separation (1.22) are also reported in the marginal 1D histograms. The 1σ Rayleigh distribution expected for the normalized separations is overplotted in green.

5. Source Characterization and Classification

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

5.1. Galactic and extragalactic sources

In order to classify sources in the most reliable way, we have used a combination of methods and various information: spectroscopic, parallax measurements from Gaia, colors and morphology from imaging surveys. None of the methods is infallible because they all depend on the quality of the data (e.g., SNR for spectra, depth and resolution of images) and because of the degeneracy in color-redshift space for many of the sources. We have therefore adopted a multi-step approach: at each step we extract from the pool of sources those that can be classified with high reliability either as *extragalactic* or *Galactic*. Figure 9 provides a graphical illustration of the decision tree adopted for the classification, together with the number of sources in each the classes. The procedure is described below in detail.

We first apply the classification based on spectroscopy or large parallax. These can be considered primary methods as they are highly pure, but certainly not complete. The sources thus classified are defined "Secure Galactic" or "Secure Extragalactic". Briefly, we define "Secure Extragalactic" all sources with spectroscopic redshift >0.002 and `NORMQ=3` (STEP 1 in Figure 9) and "Secure Galactic" all sources satisfying at least one of the criteria: 1) spectroscopic redshift <0.002 and `NORMQ=3`; significant parallax from Gaia EDR3 (above 3σ) or agreement with Hamstar counterparts (STEP 2).

Next, from the sources still in the pool, we extracted those that appear extended in the optical images. Depending on whether or not photometry from HSC is available, a source is defined as extended (EXT) if it satisfies:

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

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

$$\texttt{LS8\_TYPE} \neq \mathrm{PSF}. \quad (3)$$

The EXT sources were then flagged as "Likely Extragalactic" (STEP 3). This is considered a secondary classifier, given that, for example, in poor seeing conditions point-like sources (or stellar binary systems) would also be mis-classified as extended (see discussion presented in Hsu et al. 2014).

The sources classified as "Secure" were then projected in the LS8 g-r vs. z-$W1$ plane following Ruiz et al. (2018). There, we have empirically defined a line separator described as:

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

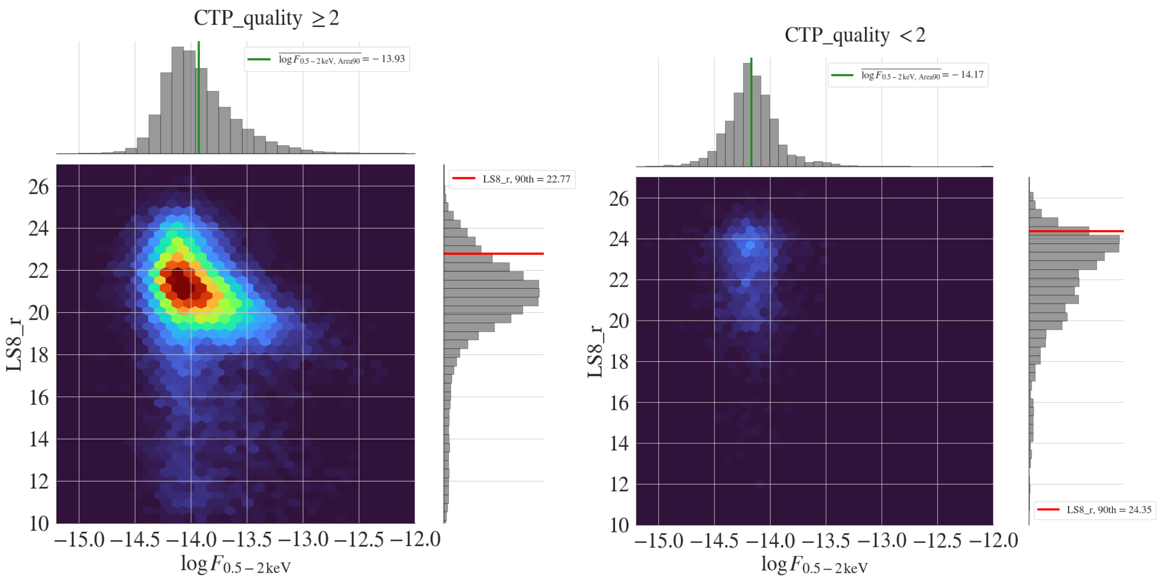


Fig. 8: Magnitude distribution of the counterpart vs the X-ray flux for sources with `CTP_quality`>= 2 (left) and `CTP_quality`< 2 (right). The magnitude distribution is clearly different. The green lines represent the mean 0.5-2 keV flux for sources marked as "inArea90" (-13.93). The red lines mark the 90-th percentile of the r-band magnitude distribution (22.77 and 24.34, respectively).

which provides a sharp separation between secure Galactic and extragalactic sources, with a negligible fraction of secure extragalactic sources lying below the separator (left panel of Figure 10). Then, for all the sources still in the pool and with available photometry from LS8 (STEP 4), we classify the sources below the line as "Likely Galactic" (STEP 5). The remaining sources in the pool with available LS8 photometry are classified as "Likely Galactic/Extragalactic" (STEP 6) depending on whether they fall below or above the line in the X-ray flux vs. *W1* plane defined in Salvato et al. (2018b):

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

with W1 in Vega system and X-ray flux in cgs. Originally, a similar line separator was introduced by Maccacaro et al. (1988) using X-ray and optical bands and over time tested at different X-ray flux depth or at different wavelength (e.g., NIR; see Civano et al. 2012). This new line separates 'X-ray bright' AGN from 'X-ray faint' stars, and was constructed combining data from the deep COSMOS *Chandra* Legacy survey (Marchesi et al. 2016) and ROSAT/2RXS (Boller et al. 2016; Salvato et al. 2018b). It can be considered a good separator only after the extended, nearby extragalactic sources are taken into account (see right panel of Figure 10). It has the advantage of generality, as the *W1* photometry and the X-ray fluxes are available virtually for all the eFEDS sources. Finally, for the sources without complete information from LS8, we assume they are extragalactic (STEP 7), unless they are below the W1-X line defined in Equation 5 (STEP 8).

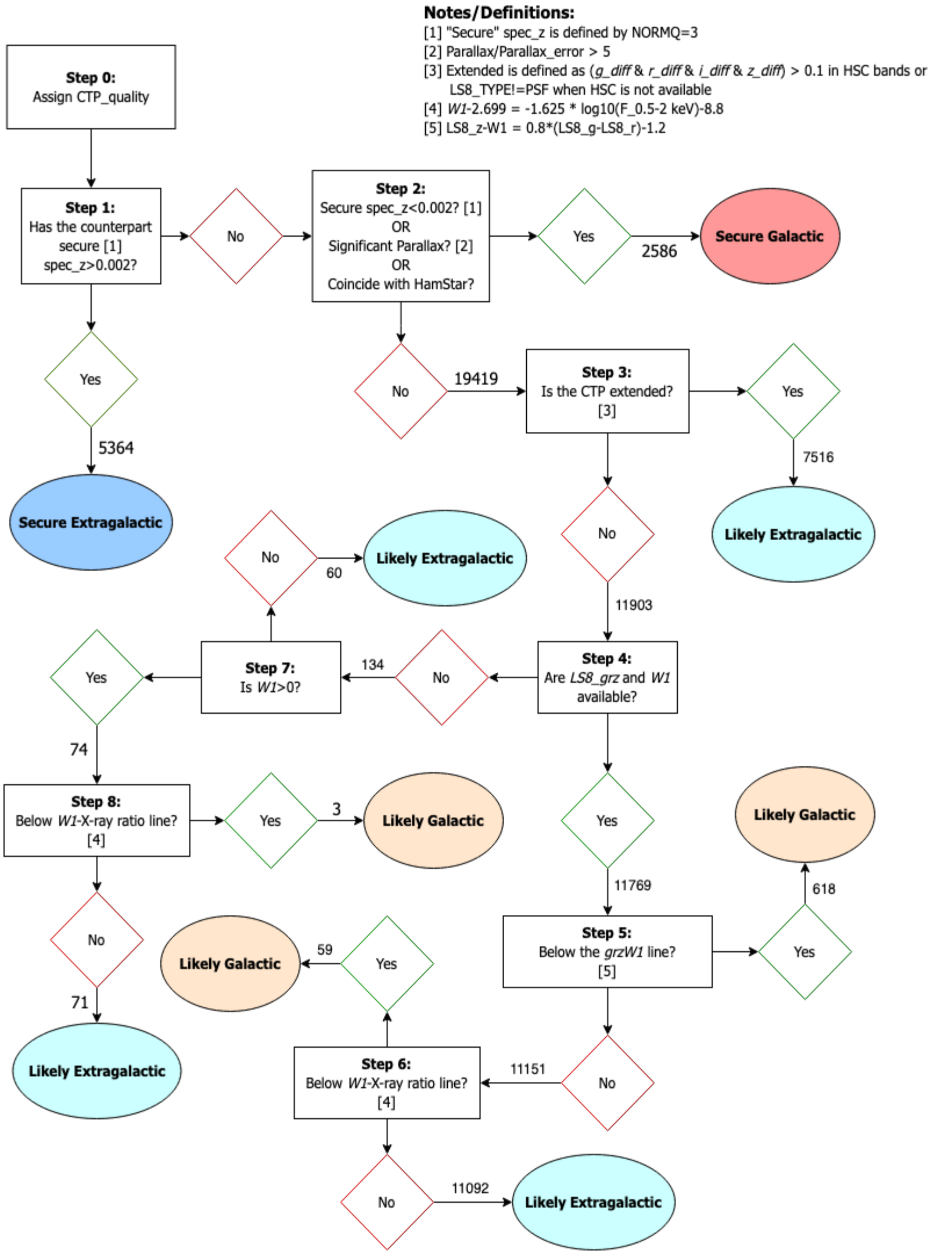
In this manner, a simple but reliable four-way classification scheme (Secure/Likely Galactic/Extragalactic) is achieved. The final distribution of the four classes of sources in the *g*-*r*-*z*-*W1* vs *W1*-*X* planes is shown in Figure 12. The two line separators identify 4 wedges that can be used for selecting sub-samples dominated by Galactic/extragalactic sources:

- Top Left: 1059 sources, out of which 965 (91.1%) are Galactic (658 and 604, respectively, only considering sources with reliable counterparts) `CTP_quality>=2`)
- Top Right: 23610 sources, out of which 23525 (99.6%) are extragalactic (21572 and 21489 for `CTP_quality>=2`)
- Bottom left: 1388 sources, out of which 1379 (99.4%) are Galactic (1317 and 1308 for `CTP_quality>=2`)
- Bottom Right: 1312 sources, out of which 455 (34.7%) are extragalactic (1227 and 382 for `CTP_quality>=2`

It is important to keep in mind that the order of the steps taken in the decision tree is crucial for limiting the misclassification of the sources as much as possible. For example, the use of spectroscopic redshift in the first step allowed the identification of the bright and nearby extragalactic sources that would have been mis-classified as Galactic, in the *W1*-*X* plane. Similarly, the adoption of the high parallax from Gaia allowed the identification for secure galactic sources that would have been mis-classified as extragalactic in the *g*-*r* vs *z*-*W1* plane.

5.2. *Validation with external samples*

We have carried out some sanity checks of the classification framework using external catalogues, whose members are expected to be either almost completely Galactic or almost completely extragalactic in nature. We consider the following 'extragalactic' dominated catalogues: the Faint Images of the Radio Sky radio component catalogue (FIRST, White et al. 1997), the Gaia–unWISE AGN candidate catalogue (GUA, Shu et al. 2019), and a sample of galaxies with significant optical extent extracted from the LegacySurvey/dr9 catalogue (LS9exgal, Dey et al. 2019a). We also match to stars having significant proper motion measurements in Gaia EDR3 (edr3pmstars, Gaia Collaboration et al. 2020), a sample likely to be almost entirely



Notes/Definitions:
[1] "Secure" spec_z is defined by NORMQ=3
[2] Parallax/Parallax_error > 5
[3] Extended is defined as (*g_diff* & *r_diff* & *i_diff* & *z_diff*) > 0.1 in HSC bands or LS8_TYPE!=PSF when HSC is not available
[4] *W1*-2.699 = -1.625 * log10(F_0.5-2 keV)-8.8
[5] LS8_z-W1 = 0.8*(LS8_g-LS8_r)-1.2

Fig. 9: Decision tree adopted for assigning each eFEDS point source to the Galactic or extragalactic classes. First we classified the sources on the basis of the most secure methods (e.g., high confidence redshift) and then proceeded with less reliable methods (e.g., based on colors) on the sources remaining in the pool, creating less pure samples. Note that the numbers listed at each step include all sources, i.e., including also those with an insecure counterpart (`CTP_quality`<2)

.

populated by galactic objects. None of these external catalogues is used directly by our classification scheme (although several are derived from common raw data sets), and so this is a semi-independent test.

GUA objects were considered when they had PROB_RF $>$ 0.8. Significantly extended sources were selected from the LS9 'sweep' files when the following criteria were satisfied: TYPE≠PSF, SHAPE_R $>$ 0.5, and SHAPE_R $\cdot$ $\sqrt{\mathrm{SHAPE\_R\_IVAR}}$ $>$ 5. Stars with significant proper motion were selected from Gaia EDR3 when $\sqrt{(pm\_ra/pm\_ra\_err)^2 + (pm\_dec/pm\_dec\_err)^2} > 5$. Simple positional matches were made against our best matching optical counterpart positions, with a search radius of 3 arcsec for the

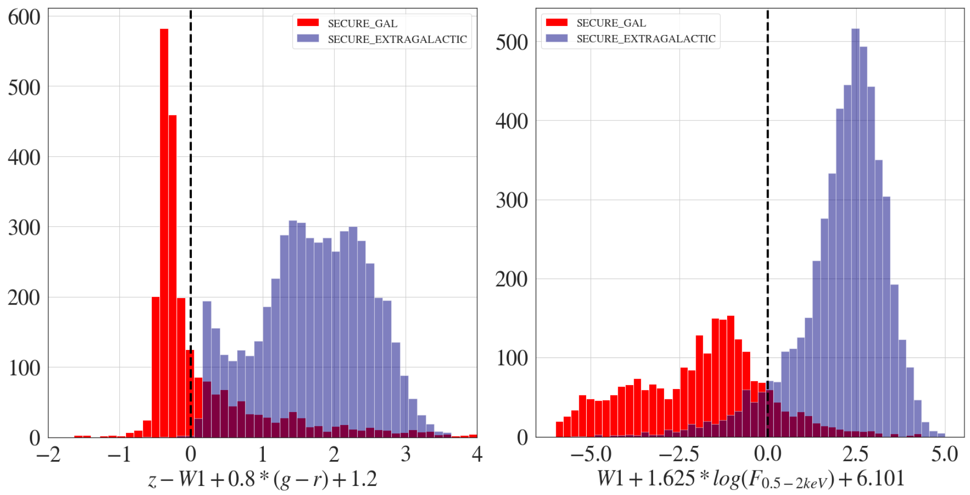


Fig. 10: The distribution of sources flagged as "Secure Galactic" (red) and "Secure Extragalactic" (blue) in the *g*-*r* vs *z*-*W1* (left) and W1 vs X-ray (right) planes are used for determining a line separator (black line) to be used for classifying sources in steps 5, 6 and 8 of the flowchart presented in Figure 9. The line separator on the right has fewer Galactic sources that fall into the extragalactic locus. However, the line separator defined on the left has literally only a handful of extragalactic sources falling into the Galactic locus, making this classifier more efficient, provided the four photometric points are available.

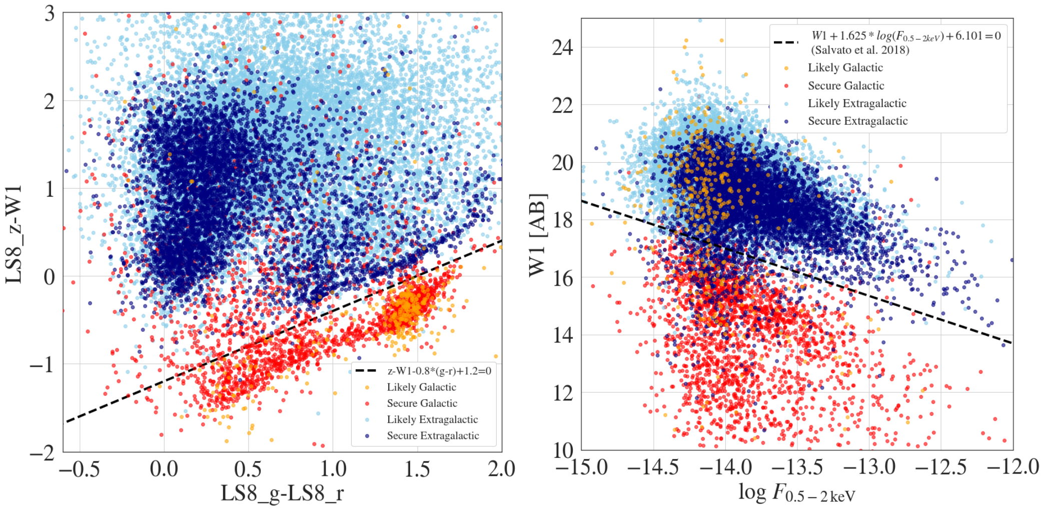


Fig. 11: The four classes "Secure Galactic" (red), "Likely Galactic" (orange), "Secure Extragalactic" (blue) and "Likely Extragalactic" (cyan) defined in the flowchart presented in Figure 9 distributed on the *g*-*r* vs *z*-*W1* (left) and W1 vs X-ray (right) planes. The black dashed lines are the empirical separators used in the classification process. See text for more details.

FIRST radio component catalogue[12] , and 1 arscec for the rest. We examine the rate at which sources we classify as Galactic or Extragalactic (both Secure and Likely) are matched to objects in these external catalogues (see Table 6). There is a very low rate of apparent disparities between our classifications and those that may be imputed by matches to the external catalogues. For example, only 0.19% of "Secure Galactic" sources have a radio counterpart in FIRST, compared to 7.1% of the "Secure Extraglactic" sample. Likewise, only 0.23% of "Secure Galactic" sources are selected by (Shu et al. 2019) as likely AGN, compared to 54% of the "Secure Extragalactic" sub-sample.

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

By construction the eFEDS X-ray point-source catalog is expected to have very little contamination from clusters of galaxies; still, a small probability that a source is actually a cluster remains, as shown from the simulations we performed for eFEDS (Liu et al. submitted). The reasons for clusters ending up in the point-like sample are manifold (Willis et al. 2021). Most obvi-

---

[12] we only considered the radio components and made no attempt to handle complex sources appropriately

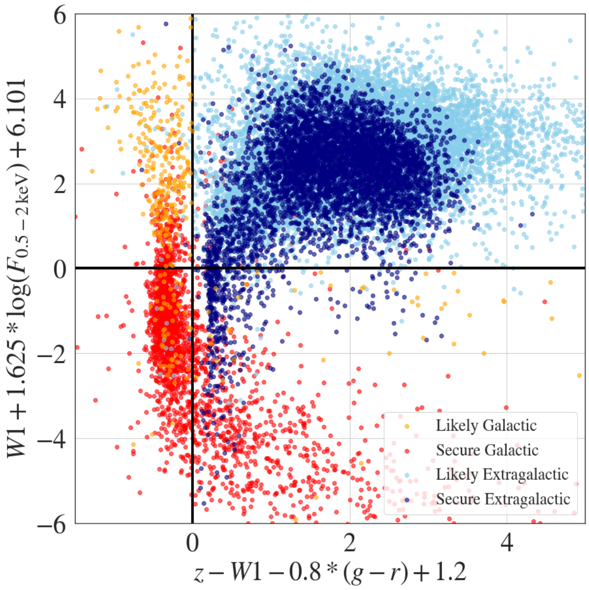


Fig. 12: The main classes of the eFEDS X-ray sources distributed according to their distance from the two lines defined in Figure 10. Three of the four wedges thus defined contain extragalactic or Galactic samples that are up to 99% pure (see text for details).

Ref. sample	Total matches	Extragalactic Secure	Extragalactic Likely	Galactic Likely	Galactic Secure
All eFEDS	27369	5364	18719	700	2586
FIRST	806	380	418	3	5
GUA	6310	2916	3381	7	6
LS9exgal	4961	1286	3657	14	4
edr3pmstars	2661	7	51	286	2317

Table 6: Comparison of our classification scheme against several semi-independent reference catalogues, suggesting that the classification is robust.

ously, clusters with small apparent size or at low detection likelihood can fall below the thresholds used to define the extension of the X-ray source. In addition, clusters could leak into the point-like sample because of source splitting and superimposition of a bright point source and a cluster.

With this in mind we have run the multi-component matched filter cluster confirmation tool (MCMF; Klein et al. 2018, 2019) on the eFEDS point source catalog. We run MCMF as in the eFEDS extended sources catalog (Klein et al., submitted), after adjusting some of the parameters (for example, limiting the area search from the X-ray position).

As in Klein et al., we can define a "contamination fraction", f_{cont}, which expresses the probability for an optical concentration of red galaxies to be a chance alignment along the line of sight to the X-ray source. This is the key selection criterion for selecting cluster candidates and immediately provides an estimate of the catalog contamination. A catalog created by selecting $f_{\mathrm{cont}} < a$ is expected to have contamination fraction of a, assuming the input catalog is highly contaminated.

Because of the high number density of sources in the point-like sample and/or the possibility that the emission from an actual cluster is split in many point sources, it can happen that many close X-ray sources point to the same optical cluster. A simple cut in f_{cont} will therefore yield a much larger sample of sources than real clusters in that catalog, causing the contamination fraction to be much higher than expected. To compensate for that, we define an "environmental flag" which is set to *true* when two additional criteria are simultaneously satisfied. Namely, the X-ray source needs to be the closest one to the associated optical structure and it needs to be more than 0.75 Mpc away from a cluster detected in the extent selected sample (Klein et al., submitted, Liu et al., submitted) at similar redshift. Only when the flag is set to *true* is the point-source further considered as a candidate for being a cluster.

In addition, following Klein et al., we have computed photo-z for the point-like sources assuming they are passive galaxies. If the quality of the fit is high then the redshift is assumed to be reliable (we refer to these redshifts as z_{red}) and it is compared with the redshift of the cluster obtained by MCMF and with the photo-z computed with LePhare (see Section 6).

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

- `Cluster_class=5`: `CTP_QUALITY`≤ 1 & $f_{\mathrm{cont}} < 0.2$ and the environmental flag set to *true*: the counterpart is unreliable and the X-ray emission is more likely associated to a cluster (top-left panel of Figure 13; 120 cases).
- `Cluster_class=4`: `CTP_QUALITY`≥ 2 & $f_{\mathrm{cont}} < 0.2$ with the environmental flag set to *true* and the redshift computed assuming galaxy templates coincides with the redshift of the optical cluster: the counterpart is reliable and the point source is a galaxy member (possibly the BCG) of the optically detected cluster (to-right panel in Figure 13; 144 cases).
- `Cluster_class=3`: `CTP_QUALITY`≥ 2 & $f_{\mathrm{cont}} < 0.2$ and the environmental flag set to *true* and the redshift computed assuming an AGN template is consistent with the redshift of the optical cluster: the counterpart is correct and the source is a cluster member (bottom-left panel from the left of Figure 13; 54 cases).
- `Cluster_class=2`: `CTP_QUALITY`≥ 2 & $f_{\mathrm{cont}} < 0.01$ and the environmental flag set to *true* while the photo-z computed by the three methods are in disagreement: the counterpart is reliable and the source (AGN) is just projected on a likely cluster (bottom-right panel of Figure 13; 39 cases).

In all the other cases the X-ray emission is from a genuine point-source and the counterpart is correct.

A dedicated effort is currently ongoing to confirm the secure clusters in the point source catalog as well as characterising and measuring their X-ray and radio properties of the confirmed clusters (Bulbul et al., in prep).

5.4. Very nearby galaxies

Unlike what happens in pencil-beam surveys, within eFEDS there are numerous very nearby and thus resolved galaxies. Vulic et al., (submitted) searched for eFEDS sources within the D25 ellipse of the sources in the Heraklion Extragalactic CATalogue (HECATE) of nearby galaxies (Kovlakas et al., submitted). For

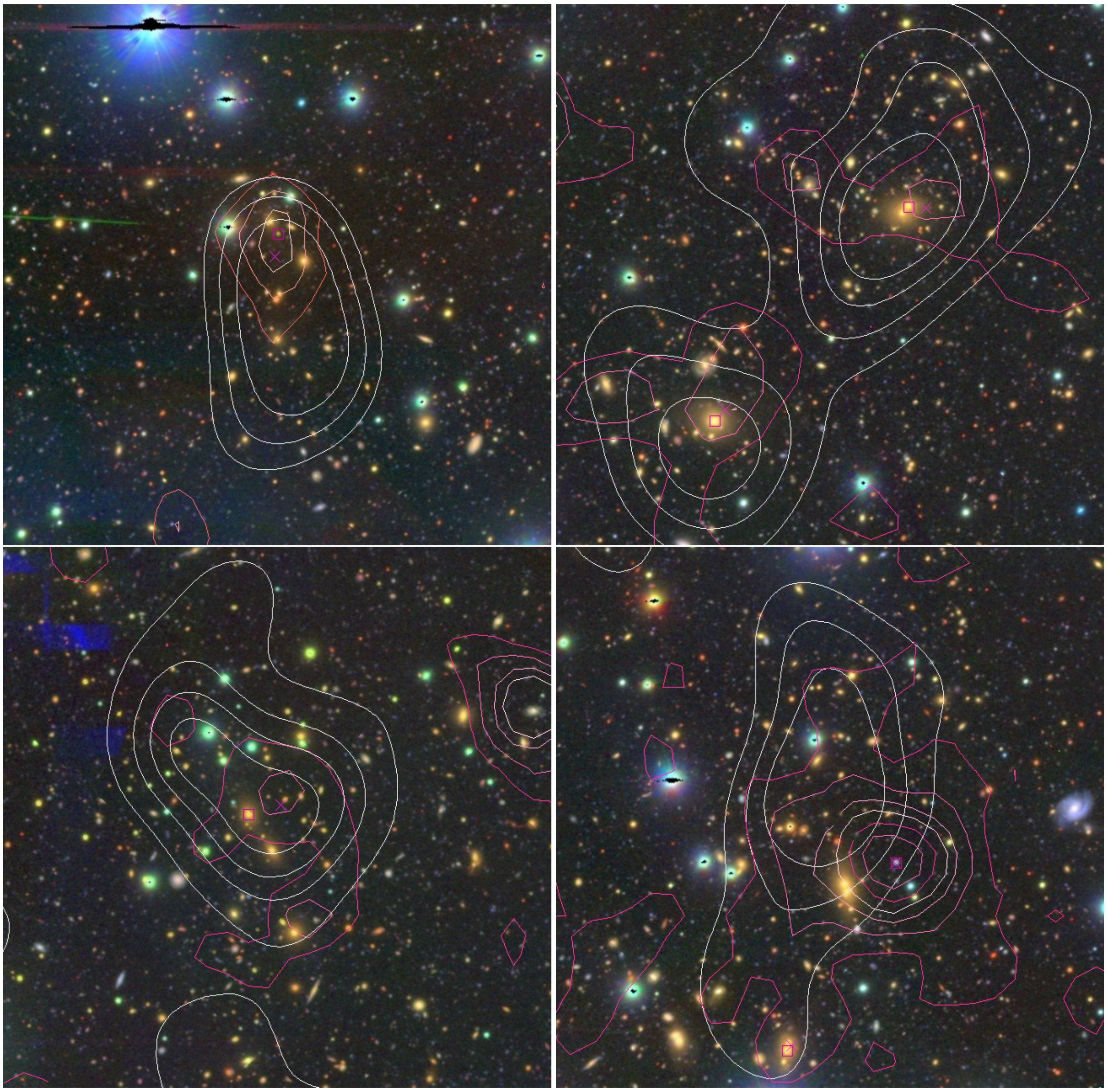


Fig. 13: from top-left,clock-wise:four examples of `Cluster_Class`=5,4,3,2, as described in Section 5.3. X-ray contours are plotted in white, with the magenta cross indicating the X-ray position and magenta square indicating the counterpart selected in this paper, magenta contours indicate the red sequence galaxy density. The HSC *g,r,i* cutouts are 5.5′×5.5′in size.

the 100 HECATE galaxies with an eFEDS source nearby, 93/100 are consistent with the counterpart proposed here by the combination of NWAY ASTROMATCH. For the remaining 7 cases (ID_SRC 7551, 12847, 2671, 22198, 17437, 29989 ,20952; see Figure .1) the counterparts identified in this work fall within the HECATE galaxies but do not coincide with the center of the galaxy but rather with a source that could be either an ULX in the galaxy or an extragalactic source in the background. For these 7 sources, dedicated studies will be needed to identify the exact origin of the X-ray emission.

6. Photometric Redshifts

Photo-z of AGN and X-ray selected sources in general have developed dramatically in the last 10 years, bringing the redshift accuracy and the fraction of outliers (usual quantities measured for assessing the quality of the photo-z) comparable to those measured for normal galaxies.

Regardless of whether photo-z are computed via SED fitting or via Machine Learning, accurate photo-z for AGN are less straightforward to be obtained than for non-active galaxies (see Salvato et al. 2018a, for a review on the topic). The main reason is that, for each multi-wavelength data point, the relative contribution of host and nuclear emission is unknown and redshift-dependent, with redshift being the parameter that we are trying

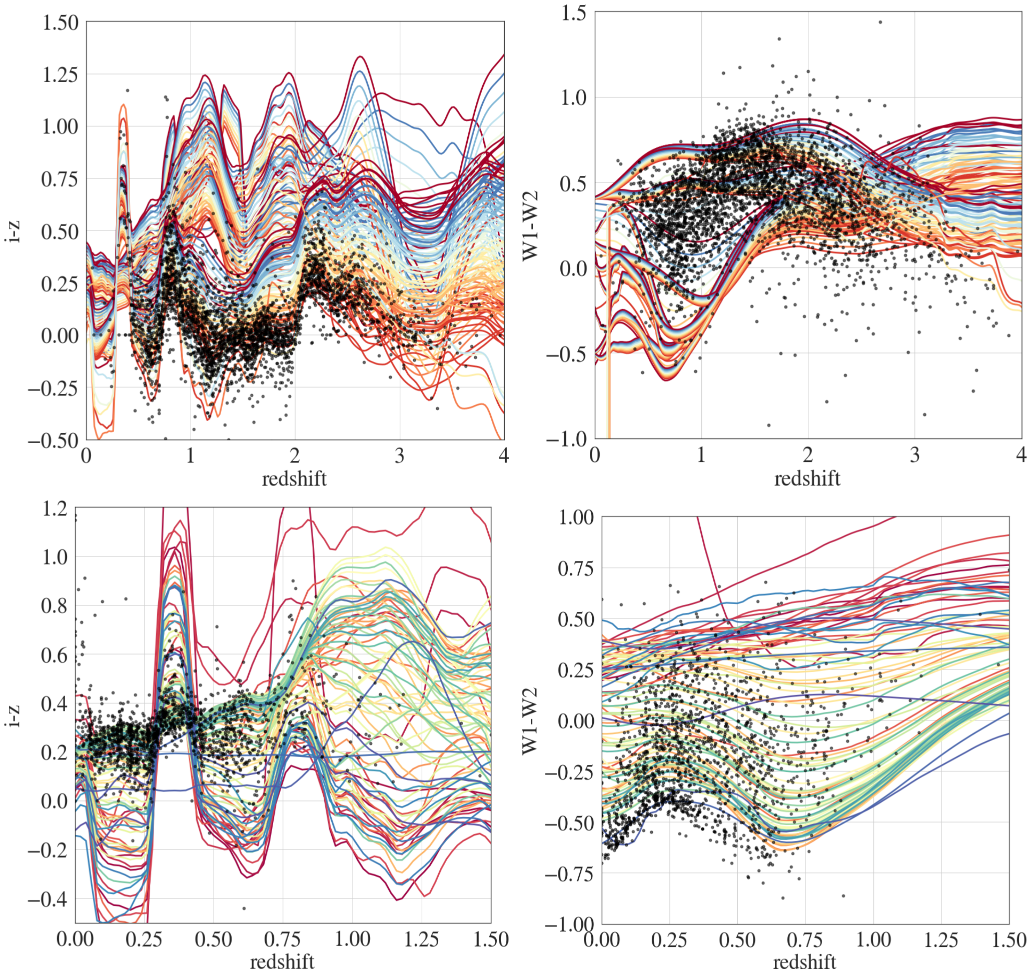


Fig. 14: i-z and W1-W2 colors of eFEDS extragalactic sources as a function of their (reliable) redshift (black points). Overplotted are the tracks of theoretical colors as a function of redshift derived from all the templates used in this work for the PLIKE (upper panels) and EXT (lower panels) samples. These kinds of plots have been used for deciding which templates should be included in the final library.

to determine. To add to the difficulty, one should not forget the impact of dust extinction and variability, the latter an intrinsic property of AGN. This, especially for wide-area surveys where data are taken over many years, can affect noticeably the accuracy of photo-z if not accounted for (e.g., Simm et al. 2015), as for example was possible to do in COSMOS (Salvato et al. 2009, 2011; Marchesi et al. 2016). In eFEDS, we also have to face the issue that the photometry is not homogenised, and different surveys cover different parts of the field at different depth and with different ways of computing the photometry (Kron, Petrosian, apertures, model etc). In the following we describe the procedure adopted for computing photo-z using LePHARE (Arnouts et al. 1999; Ilbert et al. 2006).

We computed the photo-z for the sources classified as extragalactic. In order to minimise systematic effects, we have used different types of photometry, depending on the survey; in particular, we have tried to avoid photometry derived from models for the extended and nearby sources. This is because usual models are good representation of point-like, disk-like and bulge-like sources, but are unable to represent, for example, a local Seyfert galaxy where nuclear and host components would be both contributing to the total flux. For this reason, we have used total fluxes from GALEX; Kron and cmodel photometry from HSC, depending on whether the source is extended or not (see below), and GAAP (Gaussian Aperture and Photometry) from KiDS+VIKING. From VHS we have adopted Petrosian photometry as it appears to be more in agreement with the VISTA/VIKING photometry. All the photometry was corrected for Galactic extinction, using E(B-V) from LS8. Depending on whether the source is in the area covered by KiDS+VIKING, within HSC but outside KiDS and outside HSC, different bands were available[13]

---

[13] In particular, note that for HSC, in the S19A release available to us at the time of this work, photometry in *r2* and *i2* filters is provided. However the filters have changed during the survey and depending on the coordinates of the sources the fraction of data obtained with the original or the new filters changes. In order to account for this at any location, we have adopted the filter that was used for obtaining at least

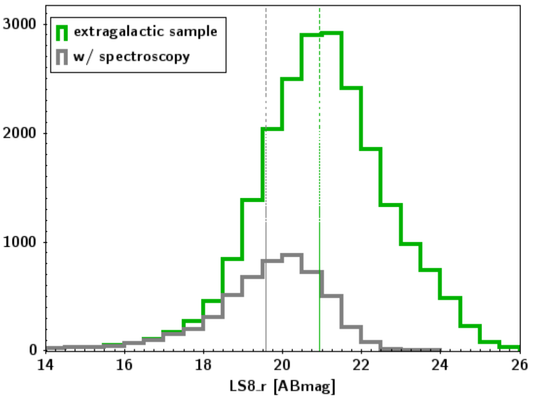


Fig. 15: Magnitude distribution for the entire extragalactic sample and the subsample for which reliable spectroscopy is available. The vertical lines indicate the median values of the two samples.

The computation of the photo-z followed the procedure already outlined in Salvato et al. (2009, 2011); Fotopoulou et al. (2012); Hsu et al. (2014); Marchesi et al. (2016); Ananna et al. (2017), where sources are treated differently, depending on whether the optical images indicate an extended (`EXT`) and thus likely nearby source, or a point-like/unresolved (`PLIKE`) source. This step is particularly important when the number of photometric points is limited and the degeneracy of the results more frequent. Within the HSC area, we classified as extended all the sources that have a difference between Kron and PSF magnitude larger than 0.1 in *g,r,i,z*, simultaneously (see Equation 2), while flagging as point-like all the others. Thus constructed, the EXT sample is not complete but is as pure as possible, although the change of the seeing may affect the measurements and thus the classification. This is a well known effect already discussed in Hsu et al. (2014), where it has been shown how up to 30% of the *Chandra* sources in the CDF survey change classification depending on whether ground- or space-based images are used for the classification. Outside HSC, we have used `TYPE` from LS8: we defined `PLIKE` the sources with `TYPE=PSF` and `EXT` all the others (see Equation 3).

The fitting templates are selected on the basis of the X-ray depth and coverage of the surveys, keeping in mind that e.g., bright AGN will be mostly absent in a deep pencil-beam survey, characterised instead by host-galaxy dominated sources. Given the similar X-ray depth, the libraries used in Ananna et al. (2017) for the Stripe-82X survey have been a good starting point for our work on eFEDS. However, recently a new library of templates for AGN and hybrids (AGN and host) was presented in Brown et al. (2019). The authors used photometry and archival spectroscopy of 41 AGN to create an additional set of 75 new hybrid templates. With respect to previous AGN templates they have the advantage that they are empirical for the entire wavelength coverage and that the contribution from the host and AGN components is fully taken into account when creating the final SED, including dust attenuation and emission lines. In addition eFEDS is particularly rich in sources with reliable spectroscopy (see Section 2.3, allowing for a better tuning of the templates to be used for the photo-z computation.

To optimize the template choice, the colors of all the sources with reliable spectroscopy were plotted as a function of redshift, together with the theoretical colors from all the templates available. Figure 14 illustrates this for *i-z* and *W1-W2* for the `EXT` and `PLIKE` samples, respectively. Various combination of templates were tested, trying to limit their number (to control degeneracy in the redshift solution), while at the same time compiling a list representative of the entire population. Important to keep in mind, though, is the fact that despite being rich, the available spectroscopic sample is not representative of the entire eFEDS population as can be seen in Figure 15. For this reason the final library should also include some templates for types of sources that are expected to be present in eFEDS, without being necessarily identified yet. In particular we have created a set of templates using the archetype of type 1 AGN from the counterparts of ROSAT/2RXS (Salvato et al. 2018b) observed within SDSS-IV/SPIDERS presented in (Comparat et al. 2020), extended in UV and MIR with various slopes. For the non empirical templates reddening was also considered, using the extinction law of Prevot (Prevot et al. 1984) with E(B-V) values from 0 to 0.4 in steps of 0.1. The selected templates are presented in Appendix A.

6.1. Reliability of photo-z

The final comparison between photo-z and spec-z, considering `EXT` and `PLIKE` sources together, for the area within KiDS+VIKING and within HSC but outside KiDS+VIKING is shown in Figure 16. We used the standard metrics for measuring the quality of photo-z (see Salvato et al. 2018a, for more details on definitions): a)the fraction of outliers η: it highlights the fraction of sources with unexpectedly large errors and it is defined as the fraction of sources for which $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec}) > 0.15$ (e.g., Hildebrandt et al. 2010). b)accuracy $\sigma_{\rm NMAD}$ that describes the expected scatter between predictions and truths and it defined as $1.48 \times \mathrm{median}(|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec}))$ (Ilbert et al. 2006).

The results are listed in Table 7. Figure 17 shows the same results, but split as a function of z-band magnitude from LS8, X-ray flux and spectroscopic redshift.

Ideally, for the best computation of photo-z, in particular for sources dominated by emission lines as AGN, photometry from broad band filters across the entire spectral range should be complemented by narrow band and near infrared photometry and should be homogenised (e.g Salvato et al. 2009, 2018a). While narrow-band photometry is not available, at least some of the surveys do provide homogenised photometry. For what concerns NIR photometry, the VISTA/VHS data are not sufficiently deep and how this impacts the photo-z is clearly visible in all the panels of Figure 17, where the fraction of outliers is usually higher and the accuracy lower (high value of $\sigma_{\rm NMAD}$ in the area without VIKING coverage (dotted lines). Not only are the NIR data shallow outside the KiDS+VIKING area; they are just a collection of photometric points computed in different ways, simply matched in coordinates. For this reason, based on the footprints shown in Figure 1, we can think of the photo-z in eFEDS as divided in three regions that reflect the quality of the available photometry: the inner area covered by deep forced photometry in KiDS+VIKING; the area that is within HSC but outside KiDS+VIKING for which some NIR information is provided by the shallow VISTA/VHS; the area outside HSC for which the optical photometry is provided only by LS8.

The lack of deep NIR data also creates an unusual number of sources at high-z (z>3), most of which are most likely incorrect. For example the number of sources with photo-z > 3 is 188 within KiDS and 819 in the HSC area outside KiDS, despite

---

50% of the data. This solution is not optimal and will affect the quality of the photo-z in some areas.

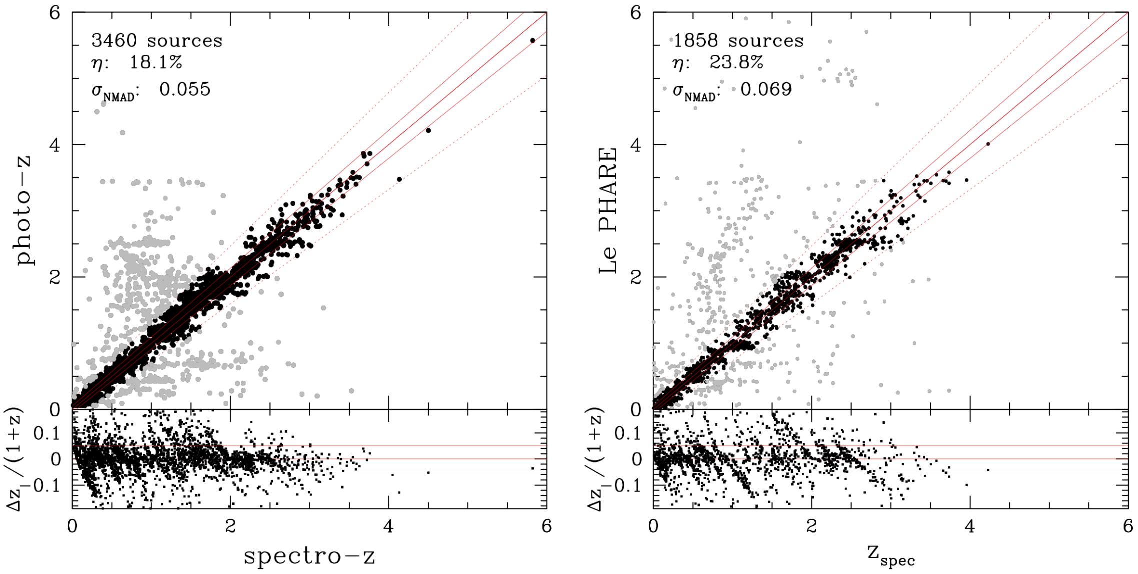


Fig. 16: spec-z vs photo-z from Le PHARE for the sources inside KiDS+VIKING (left panel) and inside HSC but outside KiDS+VIKING (right panel). The sources in gray are considered outliers while the red lines correspond to 1) z_phot = z_spec (thick solid), 2) z_phot = z_spec ±0.05 x(1+z_spec) (solid) and 3) z_phot =z_spec ± 0.15(1 + z_spec) (dotted).

AREA	sample	Le PHARE			DNNz		
		N_{spec}/N	η	$\sigma_{\rm NMAD}$	N_{spec}/N	η	$\sigma_{\rm NMAD}$
inside KiDS	PLIKE	2294/6648	17.8%	0.048	2238/6648	20.1%	0.043
	EXT	1166/4614	6.7%	0.056	1114/4622	6.4%	0.032
	TOTAL*	3460/11262	13.9%	0.049	3352/11262	15.5%	0.038
outside KiDS	PLIKE	1387/8315	29.1%	0.81	1306/8315	31.5%	0.066
	EXT	500/4292	8.8%	0.040	471/4292	5.3%	0.027
	TOTAL*	1887/12607	21.8%	0.066	1887/12607	24.6	0.048
outside HSC	PLIKE	14/146	42.9%	0.203	/N/A	N/A	N/A
	EXT	3/88	0.0%	0.039	N/A	N/A	N/A
	TOTAL	17/234	35.3%	0.113	N/A	N/A	N/A

Table 7: Fraction of outliers and accuracy for Le PHARE and DNNz computed using the extragalactic sources with secure spectroscopic redshift, split by area. The values provided for DNNz do not account for the sources for which DNNz was unable to provide a photo-z.

the area being about the same size. Within KiDS+VIKING, Le PHARE correctly estimates the redshift for 40 of the 55 (72.7%) sources spectroscopically confirmed to be at redshift larger than 3. Most of these high-z sources in excess can be easily identified and flagged by noticing that they are characterized by having high pdz[14], despite being in the area outside the HSC, i.e., with a very limited number of photometric points to be fitted.

Figure 17 also shows how the accuracy degrades and the fraction of outliers increases for the `PLIKE` that are X-ray bright (top, second panels from the left). These are sources dominated by the AGN component with an SED close to a power-law and for which the lack of narrow-band photometry that would identify the emission lines, does not allow breaking of the degeneracy in the redshift solutions. However, in eFEDS there are only 47 extragalactic sources with an X-ray flux above 5×10^{-13} erg cm^{-2} s^{-1} and a reliable spectroscopic redshift is available for 39 of them, so that the low quality of the photo-z for these sources has only a limited effect.

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

6.2. Comparison with DNNz

Within the HSC collaboration, the computation of photo-z is available in many flavors. The method that performs better on AGN is DNNz (Nishizawa et al., in prep). It is based on machine learning and uses exclusively HSC photometry, trained on the rich spectroscopic sample available for both AGN and normal

---

[14] High-redshift probability distribution pdz= $\int F(z)\,dz$ between $z_{best} \pm 0.1(1+z_{best})$ with z_{best} being the photo-z value corresponding to the best fit.

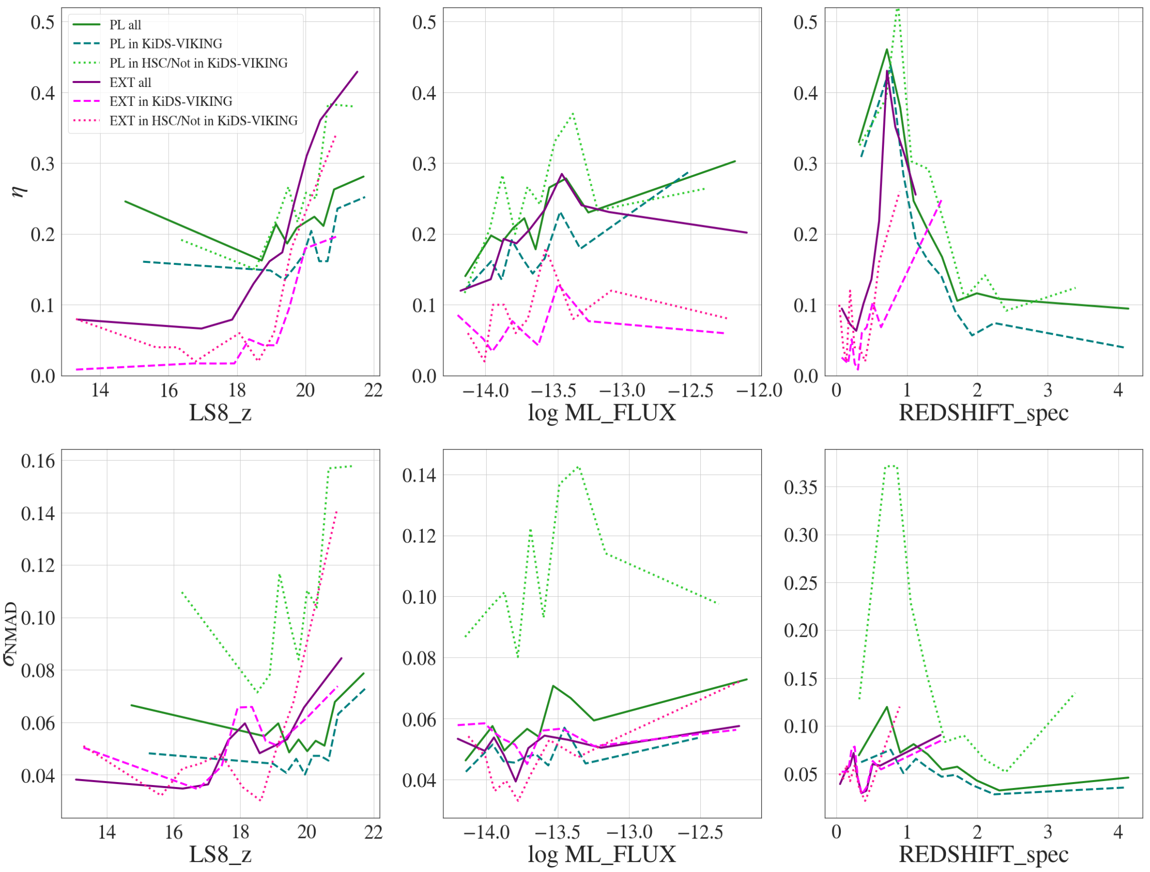


Fig. 17: Fraction of outliers (top) and accuracy (bottom) as a function of magnitude (z from LS8), 0.2-2.3 keV X-ray flux and spectroscopic redshift split for type (EXT/PLIKE) and in area (with/without VIKING coverage). The x-axes are binned in equal numbers of elements taking the quantiles between [0,1] in steps of 0.1. In other words, all lines also account for the size of each spectroscopic sub-sample.

galaxies within the entire HSC region (beyond the area in common with eFEDS). The DNNz is based on the Multi-Layer Perceptron (MLP) that takes the cmodel flux, PSF matched aperture flux, and the second order moment size measured at five HSC filter bands as inputs, and takes posterior probability as an output. In total, 3×5 inputs and output PDF is binned in 100 bins from $z = 0$ to $z = 7$. We have five hidden layers and each layer has 100 nodes where all nodes are fully connected to the nodes in the neighboring layers. With a 50k spectroscopic sample, it takes almost a single day to train this machine with NVIDIA GeForce RTX 2080Ti GPU.

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

Interestingly, despite using only HSC photometry also DNNz shows a remarkable difference in the quality of the photo-z for the sources within or outside the area covered by KiDS+VIKING. This is probably due to the combined photometry from the filters r and $r2$ and i and $i2$ that were changed during the survey. Most of the KiDS+VIKING area has been homogeneously observed only in i- and r- band, while the rest of the area has a mixture of observations. Taking this into account, we can compare Le PHARE and DNNz in the area within KiDS+VIKING and split by `TYPE`. Figure 18 shows how both sets of photo-z suffer from some systematics (vertical and horizontal substructures) due on one side to the imbalance between galaxies and AGN in the training of DNNz and on the other hand to the degeneracies in the solution for power-law dominated AGN and limited availability in photometry for Le PHARE.

However, *when the photometry is sufficient and of good quality*, SED fitting can predict correctly the redshift of AGN also when higher than 3 (middle panel), as demonstrated in Wolf et al. (2021). This is a current limitation for photo-z computed via machine learning given the small sample of this kind of sources available for training (see Nishizawa et al., submitted).

When the photo-z from DNNz is available, for each source we can measure the mean photo-z between the values proposed by the two methods. Assuming this value as the right one, we have that for 61.6% of the extragalactic sources with `CTP_quality`≥2 DNNz and Le PHARE agree ($|zp_{\rm LePHARE} - zp_{\rm DNNz}| < 0.15 \times (1 + mean(zp_{\rm LePHARE}, zp_{\rm DNNz}))$). The comparison between Le PHARE and the spectroscopic redshift for 3845 sources with spec-z is shown in the third panel of Figure 18; the

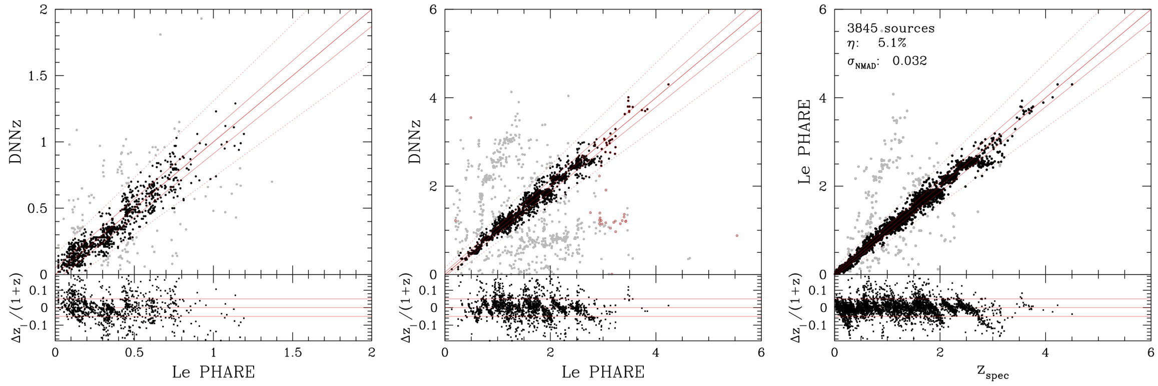


Fig. 18: Direct comparison between photo-z computed in this work with Le PHARE and DNNz, within the KiDS+VIKING area for EXT (left panel) and PLIKE (central panel). By construction true EXT sources should not have spectroscopic redshift exceeding z ≈ 1. It is not possible to decide a *a priori* whether the photo-z are wrong, or whether the sources have been erroneously put in the EXT sample due to some issue of the photometry. For the sources for which Le PHARE agrees with DNNz, the fraction of outliers is very low (right panel).

fraction of outliers with respect to the spectroscopic sample is extremely small and the accuracy very high, comparable to the accuracy routinely obtained for normal galaxies, using purely broad band photometry.

Table 8 summarizes the result for DNNz and Le PHARE separately, within and outside KiDS+VIKING. For the sources for which the two methods provide results in disagreement, the spectroscopic sample does not help in discriminating the best photo-z, given that the spectroscopic sample is very small (712 and 500 sources in the two areas, respectively) and not representative of the magnitude distribution of the about 7500 sources in the sample (mean *r* value of the spectroscopic sample 20; mean *r* value of the sample for which DNNz and Le PHARE disagree 21.5. See also next section and Figure 19). Photo-z derived via machine learning are well known to be very reliable only within the parameter space represented by the training sample, having little predictive power outside (e.g., Brescia et al. 2019). Keeping this in mind, we decided to rely on the prediction power of SED fitting and rely on the results from Le PHARE. However, we report also the results from DNNz and flag the sources for which Le PHARE and DNNz agree, or disagree, respectively (see next Section 6.3).

6.3. *`CTP_REDSHIFT` and `CTP_REDSHIFT_GRADE` in the final catalog*

In the final catalog we report the spectroscopic redshifts (regardless of their reliability) and the photo-z from both Le PHARE and DNNz. In addition, we summarise in the two columns `CTP_REDSHIFT` and `CTP_REDSHIFT_GRADE` our best knowledge of redshift and its reliability.

- `CTP_REDSHIFT`: For all sources, we use the original spectroscopic redshift when it is available and reliable (`NORMQ`=3). The redshift is set to 0 for all the sources that are classified as GALACTIC (either SECURE or LIKELY) and for which reliable redshift is not available. To the remaining sources we assign the photo-z from Le PHARE.

- `CTP_REDSHIFT_GRADE`: We assign the higher grade (5) to the sources with reliable spectroscopic redshift (6640 in total, 6519 among the sources with `CTP_quality` ≥ 2). The second higher grade (4) is assigned to the sources for which the photo-z from Le PHARE and DNNz agree (10926 in total, 9668 of which with `CTP_quality` ≥ 2), because in the previous section we have demonstrated that for this subsample the fraction of outliers is very small and the accuracy very high. By construction, all the Galactic sources without spectroscopic redshift have `CTP_REDSHIFT_GRADE`=4 because DNNz and Le PHARE are set to zero.
Ideally we would like to define the minimum number of photometric points necessary to obtain a reliable photo-z but this of course depends on the redshift of the source, its type and whether the available photometric points are those able to pinpoint key features (e.g., Ly_α, Balmer break). Thus, while we can not be sure that a source with high `pdz` has a reliable photo-z (especially at high-z), we can be sure that a low `pdz` implies low quality of the fit. We tested various thresholds for the reliability of the `pdz` (see Table 9) searching for the value that identifies the larger number of outliers in that group and also the larger number of sources at z_phot>4. The best compromise is for `pdz`<40, where still more than 70% of the sources with spec-z are outliers. Based on this criterion, we assign `CTP_REDSHIFT_GRADE` =3 to the sources with Le PHARE and DNNz in disagreement and `pdz` > 40 (6476 in total, 5885 among the sources with `CTP_quality` ≥ 2). The 1317 sources (1097 if counting only `CTP_quality` ≥ 2) with Le PHARE and DNNz in disagreement and a pdz <40 have the least reliable redshift (`CTP_REDSHIFT_GRADE`=2).

Figure 19 shows the distribution of the sources for each of the `REDSHIFT_GRADE` in the magnitude redshift plane. Indicated are also the mean value of the redshift and magnitude for each of the subsamples.

7. Data Release

The catalogs listing the properties of the counterparts to eFEDS point like sources in the Main and Hard selected samples (Brunner et. al., submitted) associated with this paper are available via CDS/Vizier and via the web page at MPE dedicated to the

Le PHARE		in KiDS	outside KiDS
	Le PHARE and DNNz agree	3.7% [98/2610]	8.7% [108/1237]
	Le PHARE and DNNz disagree	49.7% [369/742]	51.7% [279/540]
DNNz		in KiDS	outside KiDS
	Le PHARE and DNNz agree	4.0% [104/2610]	7.2% [89/1237]
	Le PHARE and DNNz disagree	55.9% [415/742]	64.4% [348/540]

Table 8: Fraction of outliers for the photo-z computed with Le PHARE and DNNz split by area, wrt the number of sources with spectroscopic redshift for which the two methods agree (top) or disagree (bottom). The small difference in the fraction of outliers for the two methods when they agree, depends on how close they are to the real spectroscopic value.

pdz threshold	N. sources	N. sources w/spec-z	N. sources w/spec-z & outliers	N. sources w/ z_phot >4
< 20	681	60/5284	47/60	104/387
< 30	1071	94/5284	70/94	200/387
< 40	1448	135/5284	97/135	231/387
< 50	1935	197/5284	125/197	253/387

Table 9: Properties distribution for extragalactic sources with pdz lower than a certain threshold. The lower the pdz, lower the quality of the fitting.

eROSITA data release[15]. The list of the columns and their description for the two samples is available in Appendix **??**. Only the basic X-ray properties are listed here (columns 1-9). For the complete list please refer to the catalogs released in Brunner et al., (submitted). After the columns reporting the key X-ray properties of the sources, columns 10-36 report the results of the counterpart (CTP) association followed by the key parameters from NWAY, ASTROMATCH and Hamstar respectively. Next (columns 36-49) we present the photometry from the recent Gaia EDR3 release, in the original photometric system, followed by all the collected photometry, corrected for extinction (columns 51-108). To be remembered is the fact that the HSC photometry from S19A in *i*- and *r*- bands has been split in *i-, i2-* and *r-, r2-* and that Kron is listed for EXT sources while cmodel is listed for PLIKE (see Section 6). Columns 109-117 list basic properties of the sources, like whether they are within KiDS, HSC etc., while columns 118-126 deal with all the information on the spectroscopy available information related to spectroscopy when available. The output parameters from Le PHARE and DNNZ are listed in columns 127-148. The columns `CTP_REDSHIFT` and `CTP_REDSHIFT_GRADE` summarise the redshift properties of the sources as discussed in Section 4.2, while the column `CLUSTER_CLASS` refers to the results presented in Section 5.3.

In addition to the catalogs, we can provide under direct request to the first author, the redshift distribution function and SED fitting of each source in the catalog. An example is shown in Figure B.1.

8. Discussion

The size and depth of the eFEDS X-ray survey, combined with ancillary data both in photometry and spectroscopy, allows us to paint a comprehensive picture of the average population of X-ray sources that contribute the bulk of the cosmic X-ray background (CXB) flux at energies < 10 keV (Gilli et al. 2007, see e.g.), both in its Galactic and extragalactic content. The identification of the optical/IR counterparts, to a high degree of completeness and reliability, as discussed here, will facilitate detailed population studies of X-ray active stars, Galactic compact objects, and AGN. Here we briefly outline the main properties of our sample by examining in detail the distributions of the X-ray sources in various color/redshift spaces.

8.1. Population studies

Figure 20 shows the distribution of all the eFEDS sources with a secure counterpart (`CTP_quality`≥2; see section 4) in four different multi-band photometric spaces, chosen for their wide applicability to large areas of the sky.

The top panel shows sources in the *g-r* vs. *z-W1* space, color-coded by their redshift. Overlaid are a few representative tracks of various classes of extra-galactic objects. Beside the clear separation between Galactic and extra-galactic objects already discussed in section 5, the X-ray points identify clear sequences of un-obscured QSO, obscured Seyferts and inactive galaxies. The inactive galaxies are best represented by the S0 and Elliptical tracks, suggesting that some of these are the sources that are associated (or confused) with a cluster. Indeed the sources indicated by a yellow circle in the top left figure, have `CLUSTER_CLASS`=3,4 indicating that they belong to a cluster and in most cases they are the BCG (see Section 5.3). These sources are best fit by the template of a passive galaxy as the spectra for those available also suggest (e.g., lack of emission lines from star formation, strong HK lines). However, some of the spectra together with the clear features from a non star-forming galaxy also reveal the presence of broad emission lines typical of AGN (see Bulbul et al., in preparation).

The top-right panel of Figure 20 shows the distribution of points in the MIR (WISE *W1*) vs soft X-ray (0.5-2 keV) plane (same as Fig. 10). X-ray bright objects above the dashed line are typically AGN, while most of the IR bright objects below the line are Galactic X-ray emitting stars, with some contamination from nearby extragalactic objects. These sources are rare, but given the size of eFEDS, they are a non-negligible number. Thus, when using this plot for other surveys one should not forget to account for the size of the survey. The larger the surveys the less efficient the line separator is.

---

[15] `https://erosita.mpe.mpg.de/edr/eROSITAObservations/Catalogues/`

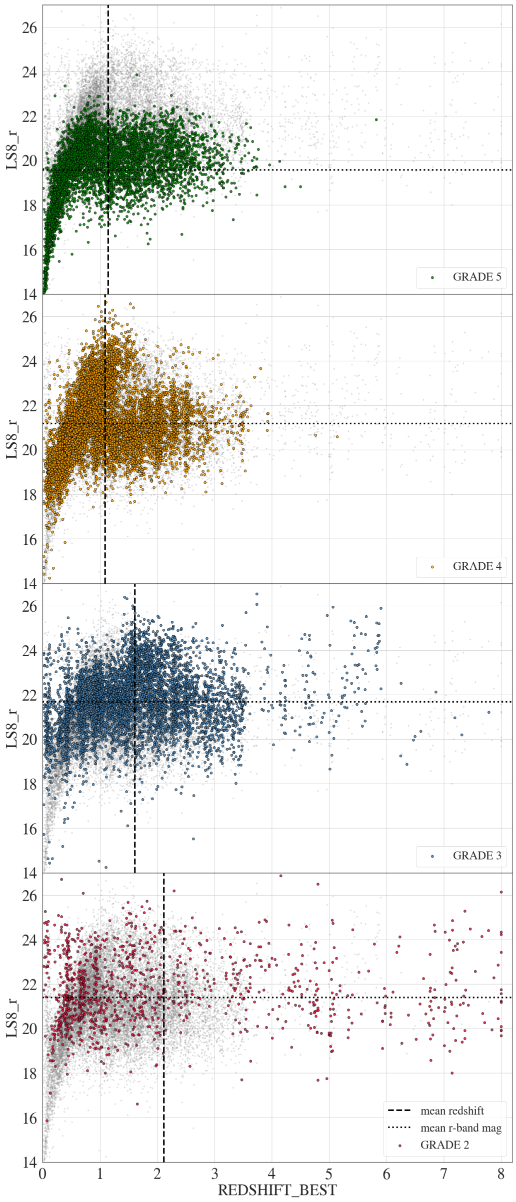


Fig. 19: eFEDS sources with reliable counterpart distributed in the magnitude vs. redshift plane, split among the various `CTP_REDSHIFT_GRADE` classes. For each panel also the mean value for redshift (vertical line) and magnitude (horizontal line) is indicated and the entire population is shown in light grey.

The bottom-right panel shows the distribution of the sources in the Wise-only *W1*-*W2* vs *W2* color-magnitude plane[16]. This is widely used to classify point sources, as it easily separates stars, with *W1*-*W2* $\approx$ 0, from QSOs, with *W1*-*W2* $>$ 0.5, (see e.g. Wright et al. 2010b; Assef et al. 2013). Once more, the eFEDS X-ray selection reveals the full extent of the extragalactic (AGN) population with intermediate IR colors between AGN- and host-galaxy dominated, typical of either obscured (Seyfert 2) or low-luminosity AGN (e.g., Merloni 2016; Hickox & Alexander 2018).

Finally, the bottom-left panel shows the distribution of the eFEDS sources in the optical/MIR diagram defined by the "all-sky available" *G*-*W1* vs *W1*-*W2*, frequently used to separate QSO from stars in the Gaia catalog. As already pointed out in Section 5, 10% of the Galaxtic sources are too faint to be detected by Gaia. This is even more true for the extragalactic sources: the plot shows only 57% of the entire eFEDS sample. However, the plot shows insights on the population that the first eROSITA All-Sky Survey (eRASS1) will uncover. As expected, the X-ray selected eFEDS sources contain, beyond stars and (unobscured) QSO, a tail at high *G*-*W1* (i.e. bright MIR, faint optical magnitudes) typical of inactive galaxies and/or mildly obscured AGN.

This is indeed confirmed by comparing the location of the extra-galactic eFEDS sources in the *grzW1* plane with the X-ray hardness ratio measured from the X-ray counts in the bands where eROSITA is most sensitive. The right-hand panel of Figure 21 shows the distribution of the sources in that plane, color-coded by their average Hardness Ratio (defined as (H-S)/(H+S) where H and S are respectively the counts in the ranges 1.0-2.0 keV and 0.2-1.0 keV[17]), while the left-hand panel highlights the loci of the most common classes of sources based on the distribution of templates tracks. Indeed, the hardest sources in the eROSITA band populate the optical/MIR color-space of Seyfert 2 galaxies and/or reddened QSO. A detailed discussion of the X-ray spectral properties of the AGN in the Main eFEDS sample will be presented in Liu, T., et al. (submitted).

9. Conclusions

We have presented the identification of the counterparts to the point sources in eFEDS presented in the Main and Hard catalogues presented in Brunner et al., (submitted) and a study of their multiwavelength properties. eFEDS, having a limiting flux of $F_{0.5-2\,\mathrm{keV}} \sim 7 \times 10^{-15}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ is a factor of $\sim$ 50% deeper than the final eROSITA all-sky survey, and because of that can be used also as a forecast for eRASS:8, not only for the population that eRASS:8 will reveal, but also for the challenges that are ahead of us with respect to counterpart identification and redshift determination.

- **Counterpart identification**: The completeness and purity in the counterpart identification reached in eFEDS is estimated to be very high (above 96%, based on a validation sample of about 3500 *Chandra* made eROSITA-like). The results, well above the predicted completeness and purity mentioned in Merloni et al. (2012) are due to three important factors: the development of new methods for identifying the correct counterparts, large samples of X-ray detected sources with known counterparts, and the availability of sufficiently deep, homogenised, multi-wavelength photometry from optical to MIR over very wide areas from which to construct the SED of these sources to be used as training. In the next two years, by the time that eROSITA has completed the final survey, the methods will continue to improve and the training/validation samples will increase in size. Most importantly, the coverage

---

[16] For this plot we are using Vega System, so that user can compare the figure with similar ones done using AllWISE all sky

[17] The hardness ratio is calculated using the columns ML_CTS_b1, ML_CTS_b2 and ML_CTS_b3 in Brunner et al. 2021 catalog: `(ML_CTS_b3 - (ML_CTS_b1 + ML_CTS_b2))/(ML_CTS_b3 + (ML_CTS_b1 + ML_CTS_b2))`

of the multiwavelength catalog usable for the identification of the counterparts will be larger. In fact, while the DESI Legacy Imaging Survey DR9 (LS9; Dey et al. 2019b), just became publicly available, the work on DR10 has started. The survey will cover most of the eROSITA-DE area of the sky at sufficient depth, thanks to the inclusion the DECam data taken via the DeROSITAS survey (PI A. Zenteno). We predict that the identification of the counterparts for the entire eRASS will be at least of the same quality of eFEDS, also in the Galactic plane, thanks to the inclusion of Gaia EDR3 recently released.

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

Acknowledgements. M.S. thanks Olivier Ilbert for the valuable feedback provided on the computation of the photo-z. J.W. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 -390783311. M.K. acknowledges support by DFG grant KR 3338/4-1. B.M. acknowledges funding from European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 860744. PCS acknowledges support by DLR grants 50 OR 1901, 50 OR 2102.

This work is based on data from eROSITA, the soft X-ray instrument aboard SRG, a joint Russian-German science mission supported by the Russian Space Agency (Roskosmos), in the interests of the Russian Academy of Sciences represented by its Space Research Institute (IKI), and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The SRG spacecraft was built by Lavochkin Association (NPOL) and its subcontractors, and is operated by NPOL with support from the Max Planck Institute for Extraterrestrial Physics (MPE).

The development and construction of the eROSITA X-ray instrument was led by MPE, with contributions from the Dr. Karl Remeis Observatory Bamberg & ECAP (FAU Erlangen-Nuernberg), the University of Hamburg Observatory, the Leibniz Institute for Astrophysics Potsdam (AIP), and the Institute for Astronomy and Astrophysics of the University of Tübingen, with the support of DLR and the Max Planck Society. The Argelander Institute for Astronomy of the University of Bonn and the Ludwig Maximilians Universität Munich also participated in the science preparation for eROSITA.

We have made use of TOPCAT and STILTS (Taylor 2005, 2006)

Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples).

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center (ADC) at National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), National Astronomical Observatory of Japan.

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID 2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID 2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID 2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Dúag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

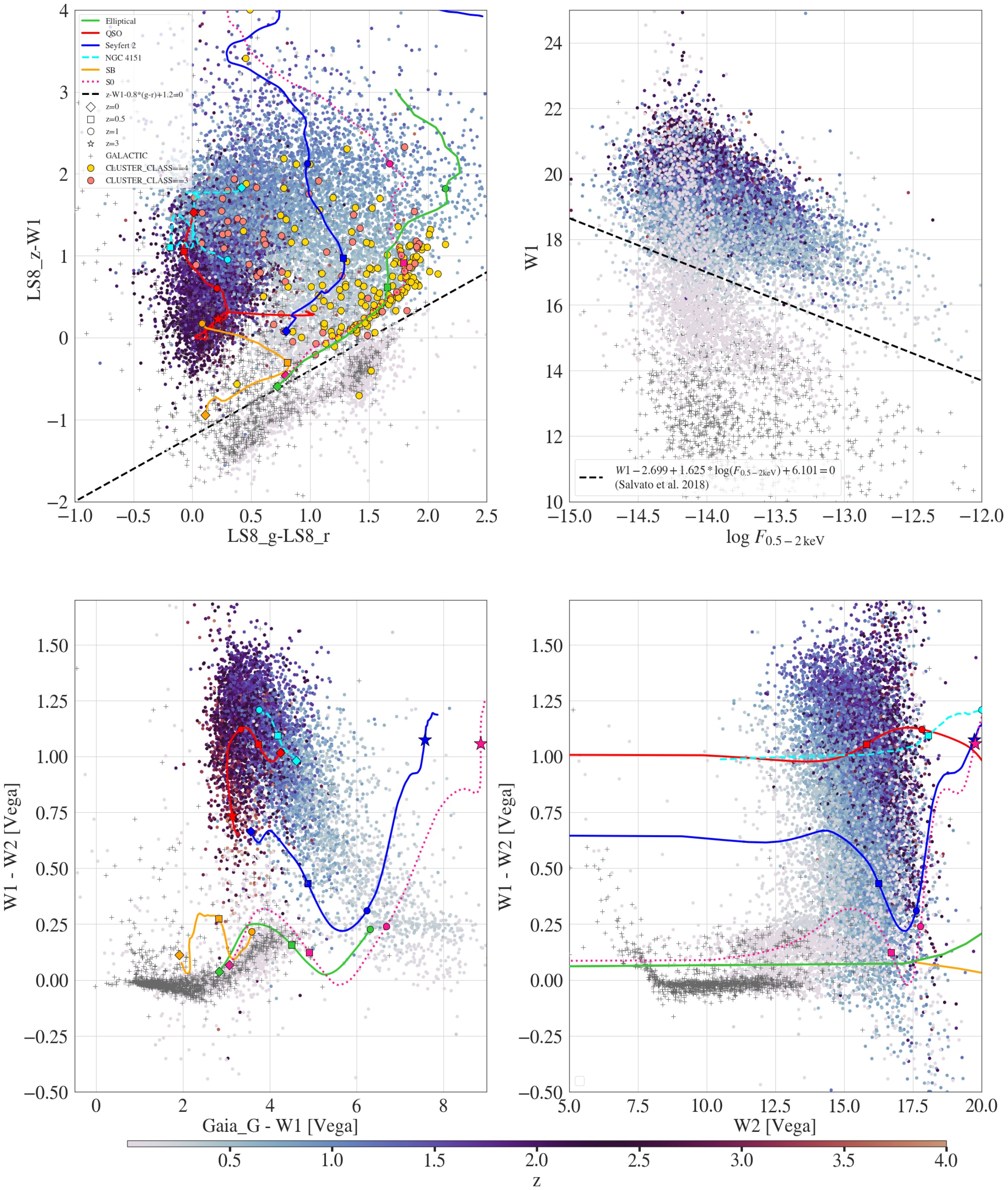


Fig. 20: The distribution of Galactic and extragalactic sources in eFEDS, color coded by redshift in the parameter spaces defined by typical colors and fluxes. For the plots with optical and MIR colors, the track of templates characteristic of the population, also used in the computation of the photo-z are overplotted. The legenda on the top-left panel provides all the details for the four panels.

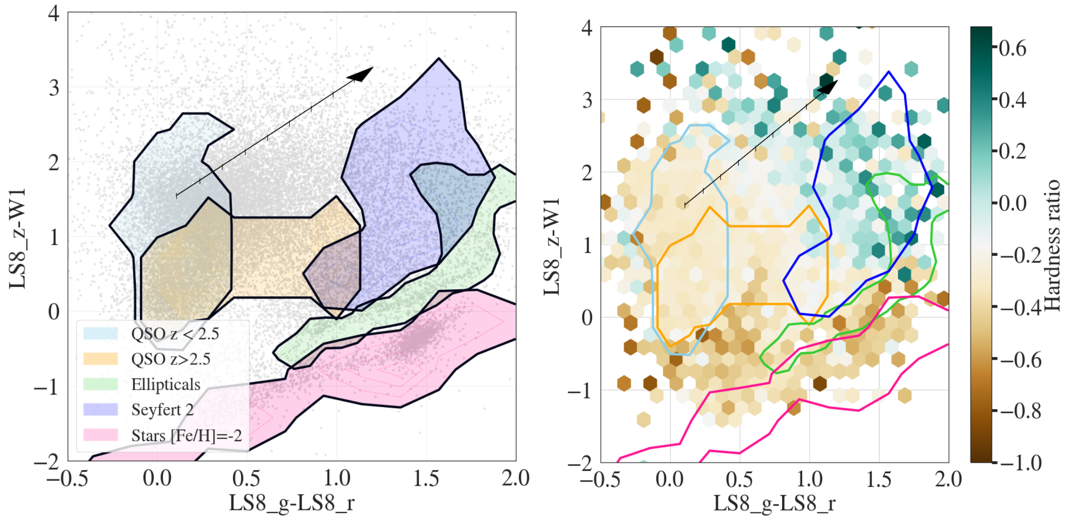


Fig. 21: *Left*: Classes of X-ray emitting sources are shown in colour-colour space. The clouds for QSOs, elliptical galaxies and Seyfert 2 galaxies are derived from theoretical colour-redshift tracks computed with `LePhare` and a subset of SED templates used for the determination of photo-z in this work. Seyfert 2 and elliptical tracks were limited to z<1, while the QSO cloud is divided in z<2.5 and 2.5<z<4. The black arrow indicates the evolution of a theoretical QSO at z=0.2 in colour-colour space with increasing extinction. The ticks along this arrow are steps of $\Delta_{E(B-V)} = 0.1$. The stellar cloud was derived from a 2D-histogram of MIST/MESA isochrones assuming [Fe/H]=-2 (Choi et al. 2016; Dotter 2016; Paxton et al. 2018). *Right:* The same classes are shown on top of the binned eFEDS data (CTP_quality $\geq$ 2), colour-coded according to the average hardness-ratio (see text for details).

Appendix A: Templates used for photo-z

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

Appendix A.1: templates used for PLIKE

The library of templates used for the PLIKE sample includes a mixture of SEDs from single objects and hybrids (different relative contribution of host and nuclear component, as introduced in Salvato et al. (2009). Templates 1 and 12-17 and 20 are originally from (Polletta et al. 2007). However, as templates N. 18 and N. 19 (originally from SDSS-V[18], template N.20 has been extended in the UV and presented in Salvato et al. (2009). The same template was then used for creating hybrids templates by mixing it with normal galaxies templates (Noll et al. 2004) with different degree of star formation, as presented in the Ananna et al. (2017)[19]. With a similar procedure the hybrid templates 2-10 where constructed combining an S0 and a QSO2 template, both from Polletta et al. (2007). The hybrids were originally presented in Salvato et al. (2009) and were successfully used in Salvato et al. (2009, 2011); Marchesi et al. (2016) among others. Finally, templates 24-29 are from the recent work of Brown et al. (2019).

1. S0
2. S0_10_QSO2_90
3. S0_20_QSO2_80
4. S0_30_QSO2_70
5. S0_40_QSO2_60
6. S0_50_QSO2_50
7. S0_60_QSO2_40
8. S0_70_QSO2_30
9. S0_80_QSO2_20
10. S0_90_QSO2_10
11. CB1_0_LOIII4
12. Sb
13. Spi4
14. M82
15. I22491
16. Sey18
17. Sey2
18. pl_QSOH
19. pl_QSO_DR2_029_t0
20. pl_TQSO1
21. s250_10_pl_TQSO1_90
22. s180_30_pl_TQSO1_70
23. s800_20_pl_TQSO1_80
24. 3C120
25. MRK110
26. NGC5548_64.00_NGC4138
27. PG0052p251
28. NGC4151
29. NGC3783_00.50_NGC4725

Appendix A.2: templates used for EXT

The templates used for computing the photo-z for the EXT sample are almost entirely taken from Brown et al. (2019). The list of AGN templates includes SED of single objects (the name of the template is self explanatory) and hybrids constructed combining in different ratio the SED of two different galaxies and AGN. Additionally we have added two templates of elliptical galaxies from Polletta et al. (2007) and two newly created templates of type 1 AGN. We have used the Type 1 archetype presented in Comparat et al. (2020) and obtained by the stacking of all type 1 sources that are counterparts to ROSAT/2RXS (Boller et al. 2016; Salvato et al. 2018b) that had an SDSS spectra. The SED was then extended in the MIR using the BQSO template presented in Polletta et al. (2007). BQSO is similar to TQSO but with less enhanced MIR flux. This new hybrid was then extended in the UV with various slopes.

1. 2MASXJ13000533+1632151
2. Ark564
3. F16156+0146
4. F2M1113+1244
5. Fairall9
6. H1821+643
7. IRAS_11119+3257
8. Mrk110
9. Mrk1502
10. Mrk231
11. Mrk290
12. Mrk493
13. Mrk590
14. Mrk817
15. NGC3227_Central_00.50_NGC4569
16. NGC3227_Central_01.00_NGC4569
17. NGC3227_Central_02.00_NGC4569
18. NGC3227_Central_04.00_NGC4569
19. NGC3227_Central_08.00_NGC4569
20. NGC3227_Central_16.00_NGC4569
21. NGC3227_Central_32.00_NGC4569
22. NGC3227_Central_64.00_NGC4569
23. NGC3516_Central_00.50_NGC4826
24. NGC3516_Central_01.00_NGC4826
25. NGC3516_Central_02.00_NGC4826
26. NGC3516_Central_04.00_NGC4826
27. NGC3516_Central_08.00_NGC4826
28. NGC3516_Central_16.00_NGC4826
29. NGC3516_Central_32.00_NGC4826
30. NGC3516_Central_64.00_NGC4826
31. NGC3516_Central
32. NGC3783_Central_01.00_NGC4725
33. NGC3783_Central_02.00_NGC4725
34. NGC3783_Central_04.00_NGC4725
35. NGC3783_Central_08.00_NGC4725
36. NGC3783_Central_16.00_NGC4725
37. NGC3783_Central_32.00_NGC4725
38. NGC3783_Central_64.00_NGC4725
39. NGC4051_Central_00.50_NGC3310
40. NGC4051_Central_00.50_NGC4125
41. NGC4051_Central_00.50_NGC4559
42. NGC4051_Central_01.00_NGC3310
43. NGC4051_Central_01.00_NGC4125
44. NGC4051_Central_01.00_NGC4559
45. NGC4051_Central_02.00_NGC3310
46. NGC4051_Central_02.00_NGC4125
47. NGC4051_Central_02.00_NGC4559
48. NGC4051_Central_04.00_NGC3310

---

[18] `http://classic.sdss.org/dr5/algorithms/spectemplates`

[19] Note that the templates are slightly different than in Ananna et al. in the UV part

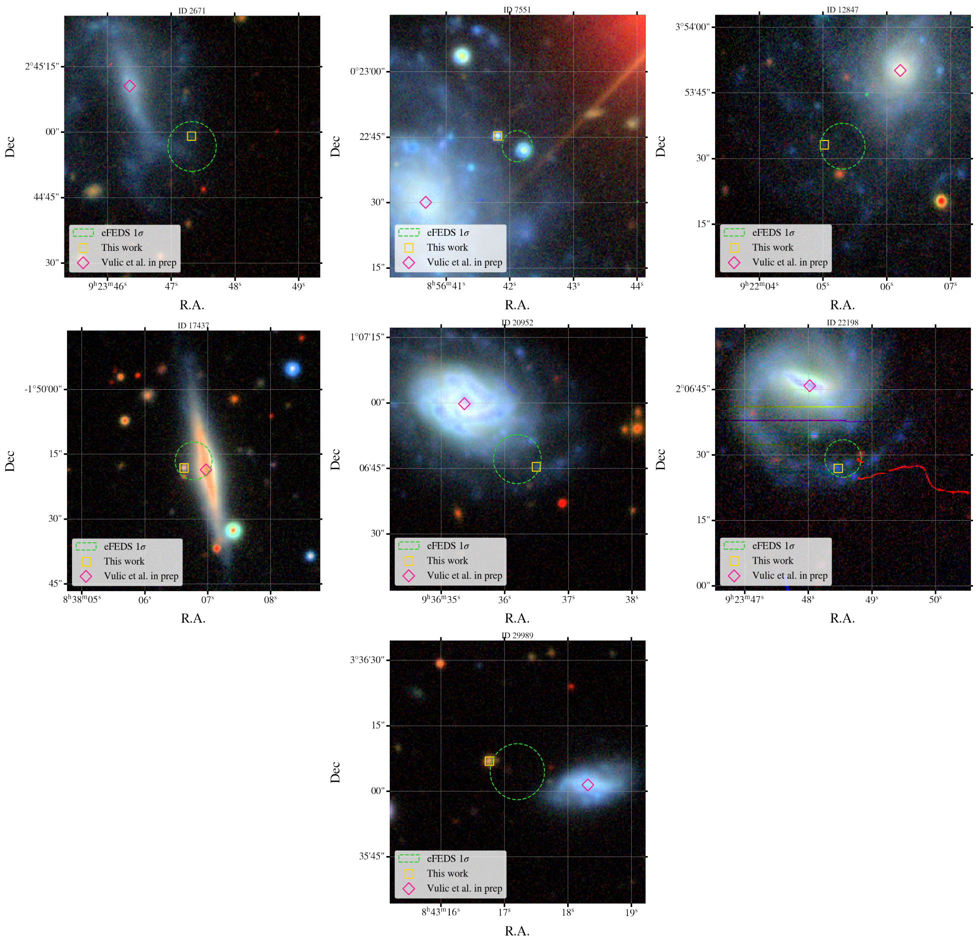


Fig. .1: The seven sources for which our counterpart fall within an HECATE galaxy (see Vulic et al., submitted) but is probably pointing to a background extragalactic source. The RGB images are 1′×1′. In the cutouts the position of our proposed counterpart, the center of the galaxy, the X-ray position and the positional error are indicated.

49. NGC4051_Central_04.00_NGC4125
50. NGC4051_Central_04.00_NGC4559
51. NGC4051_Central_08.00_NGC3310
52. NGC4051_Central_08.00_NGC4125
53. NGC4051_Central_08.00_NGC4559
54. NGC4051_Central_16.00_NGC3310
55. NGC4051_Central_16.00_NGC4125
56. NGC4051_Central_16.00_NGC4559
57. NGC4051_Central_32.00_NGC3310
58. NGC4051_Central_32.00_NGC4125
59. NGC4051_Central_32.00_NGC4559
60. NGC4051_Central_64.00_NGC3310
61. NGC4051_Central_64.00_NGC4125
62. NGC4051_Central_64.00_NGC4559
63. NGC4051_Central
64. NGC4151_Central_00.50_NGC4125
65. NGC4151_Central_00.50_NGC4579
66. NGC4151_Central_01.00_NGC3310
67. NGC4151_Central_01.00_NGC4125
68. NGC4151_Central_01.00_NGC4579
69. NGC4151_Central_02.00_NGC3310
70. NGC4151_Central_02.00_NGC4125
71. NGC4151_Central_02.00_NGC4579
72. NGC4151_Central_04.00_NGC3310
73. NGC4151_Central_04.00_NGC4125
74. NGC4151_Central_04.00_NGC4579
75. NGC4151_Central_08.00_NGC3310
76. NGC4151_Central_08.00_NGC4125
77. NGC4151_Central_08.00_NGC4579
78. NGC4151_Central_16.00_NGC3310
79. NGC4151_Central_16.00_NGC4125
80. NGC4151_Central_16.00_NGC4579
81. NGC4151_Central_32.00_NGC3310
82. NGC4151_Central_32.00_NGC4125
83. NGC4151_Central_32.00_NGC4579
84. NGC4151_Central_64.00_NGC3310
85. NGC4151_Central_64.00_NGC4125
86. NGC4151_Central_64.00_NGC4579
87. NGC5548_Central_00.50_NGC4138
88. NGC5548_Central_01.00_NGC4138
89. NGC5548_Central_02.00_NGC4138
90. NGC5548_Central_04.00_NGC4138
91. NGC5548_Central_08.00_NGC4138
92. NGC5548_Central_16.00_NGC4138

93. NGC5548_Central_32.00_NGC4138
94. NGC5548_Central_64.00_NGC4138
95. NGC5548_Central
96. NGC5728
97. NGC7469
98. OQ_530
99. PG0026+129
100. PG1415+451
101. PKS1345+12
102. Ton951
103. Ell4_A_0
104. Ell5_A_0
105. pl_BQSO_Co19_sl-8
106. pl_BQSO_Co19_sl-20

Appendix B: Release of PDZ and SED fitting

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

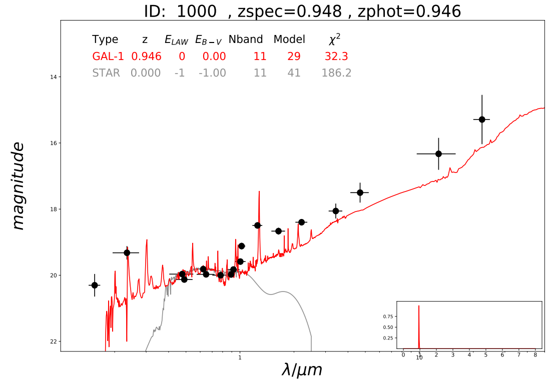


Fig. B.1: Example of SED fitting and redshift probability distribution (in the inset). The photometric points and their errors are indicated with black points. The best extragalactic and Galactic templates are shown with a solid line. The plot and the data for creating the plots are available under request.

Most columns are common to both the Main and Hard sample. We list column descriptions for the Main sample. In the few cases where there are differences in the column descriptions, we report in brackets the corresponding description for the Hard sample. The last column of the catalog (ID_MAIN) is present only in the Hard catalog.

1. **ERO_Name**: eROSITA official source Name (see Brunner et al.,)
2. **ERO_ID**: ID of eROSITA source from V18C catalog [ID of eROSITA source from V18C catalog] **(here we need to put the final name of the catalog and reference).**
3. **ERO_RA_CORR**; J2000 Right Ascension of the eROSITA source (corrected) in degrees from the V18C catalog. [J2000 Right Ascension of the eROSITA source (corrected) in degrees from the V18T catalog].
4. **ERO_Dec_CORR**: J2000 Declination of the eROSITA source (corrected) in degrees from the V18C catalog. [J2000 Declination of the eROSITA source (corrected) in degrees from the V18t catalog.]
5. **ERO_RADEC_ERR_CORR**: eROSITA positional uncertainty (corrected) in arcsec from the V18C catalog. [eROSITA positional uncertainty (corrected) in arcsec from the V18T catalog.]
6. **ERO_ML_FLUX**: 0.2-2.3 keV source flux in erg cm^{-2} s^{-1}, converted from count rate assuming ECF=1.074e+12 (Gamma=2.0). [In the Hard catalog the column is called **ERO_ML_FLUX_3**: 2.3-5 keV source flux in erg cm^{-2} s^{-1}, converted from count rate assuming ECF=1.147e+11 (Gamma=2.0)]. See Brunner et al.
7. **ERO_ML_FLUX_ERR**: 0.2-2.3 keV source flux error (1 sigma) in erg cm^{-2} s^{-1}. [In the Hard catalog the column is called **ERO_ML_FLUX_ERR_3**: 2.3-5 keV error (1 sigma) in erg cm^{-2} s^{-1}]
8. **ERO_DET_LIKE**: 0.2-2.3 keV detection likelihood measured by PSF-fitting. [In the Hard catalog the column is called **ERO_DET_LIKE_3**: 2.3-5 keV detection likelihood measured by PSF-fitting.]
9. **ERO_inArea90**: true if in the 0.2-2.3keV exp>500s region, which comprises 90%
10. **CTP_LS8_UNIQUE_OBJID:** LS8 unique identifier for the proposed counterpart to the eROSITA source (Expression: toString(LS8_BRICKID)+"_"+toString(LS8_OBJID))
11. **CTP_LS8_RA**: J2000 Right Ascension of the LS8 counterpart in degrees
12. **CTP_LS8_Dec**: J2000 Declination of the LS8 counterpart in degrees
13. **Dist_CTP_LS8_ERO**: Separation between selected counterpart and eROSITA position in arcsec.
14. **CTP_NWAY_LS8_UNIQUE_OBJID**: LS8 unique identifier of the LS8 counterpart from NWAY.
15. **CTP_NWAY_LS8_RA**: J2000 Right Ascension of the LS8 counterpart from NWAY in degrees.
16. **CTP_NWAY_LS8_Dec**: J2000 Declination of the LS8 counterpart from NWAY in degrees.
17. **CTP_NWAY_dist_bayesfactor**: Logarithm of ratio between prior and posterior, from separation, positional error and number density (see Appx. in Salvato et al 2018 for clarifications)
18. **CTP_NWAY_dist_post**: Distance posterior probability comparing this association vs. no association (see Appx. in Salvato et al 2018 for clarifications)
19. **CTP_NWAY_p_single**: Same as dist_post, but weighted by the prior (see Appx. in Salvato et al 2018 for clarifications)
20. **CTP_NWAY_p_any**: For each entry in the V18C catalogue, the probability that there is a counterpart (see Appx. in Salvato et al 2018 for clarifications)
21. **CTP_NWAY_p_i**: Relative probability of the eROSITA/LS8 match (see Appx. in Salvato et al 2018 for clarifications)
22. **Dist_NWAY_LS8_ERO**: Separation between the eROSITA position and the LS8 counterpart from NWAY in arcsec
23. **CTP_MLR_LS8_UNIQUE_OBJID**: LS8 unique identifier of the LS8 counterpart from ASTROMATCH
24. **CTP_MLR_LS8_RA** : J2000 Right Ascension of LS8 counterpart from ASTROMATCH in degrees
25. **CTP_MLR_LS8_Dec**: J2000 Declination of LS8 counterpart from ASTROMATCH in degrees
26. **CTP_MLR_LR_BEST**: Likelihood Ratio value from ASTROMATCH
27. **CTP_MLR_REL_BEST**: Reliability of the identification from ASTROMATCH
28. **Dist_MLR_LS8_ERO**: Separation between the eROSITA position and the LS8 counterpart from ASTROMATCH in arcsec
29. **CTP_SAME**: Comparison NWAY/MLR: true if the counterpart selected by the two method is the same

30. **CTP_MLR**: Comparison NWAY/MLR: true if the counterpart from NWAY(MLR) has p_any(LR_BEST) below(above) threshold
31. **Dist_CTP_NWAY_MLR**: Separation between the counterpart selected by NWAY and MLR in arcsec
32. **CTP_SECONDARY**: true if alternative counterpart exists
33. **CTP_Hamstar**: Match to Hamstar: 1=same counterpart, 0=different counterpart, -99=no Hamstar (Schneider et al)
34. **CTP_Hamstar_p_stellar**: Probability of association from Hamstar.
35. **Dist_CTP_Hamstar**: Separation between the counterpart proposed by Hamstar and the counterpart selected in this work.
36. **CTP_quality**: counterpart quality: 4=best, 3=good, 2=with secondary, 1/0=unreliable. (see flow chart paper)
37. **GaiaEDR3_ID**: ID in Gaia EDR3 source catalog
38. **GaiaEDR3_parallax**: mas Parallax from Gaia EDR3
39. **GaiaEDR3_parallax_error**: Parallax error from Gaia EDR3 in mas
40. **GaiaEDR3_parallax_over_error**: Parallax/Parallax error. ratio >5 SECURE GALACTIC
41. **GaiaEDR3_pmra**: Proper motion in RA from Gaia EDR3
42. **GaiaEDR3_pmra_error**: Error on Proper motion in RA from Gaia EDR3
43. **GaiaEDR3_pmdec**: Proper motion in Dec from Gaia EDR3
44. **GaiaEDR3_pmdec_error**: Error on Proper motion in Dec from Gaia EDR3
45. **GaiaEDR3_phot_g_mean_mag**: g band magnitude (VEGA) from Gaia EDR3
46. **GaiaEDR3_phot_g_mean_mag_error**: Error g band magnitude (VEGA) from Gaia EDR3
47. **GaiaEDR3_phot_bp_mean_mag**: bp band magnitude (VEGA) from Gaia EDR3
48. **GaiaEDR3_phot_bp_mean_mag_error**: Error bp band magnitude (VEGA) from Gaia EDR3
49. **GaiaEDR3_phot_rp_mean_mag**: rp band magnitude (VEGA) from Gaia EDR3
50. **GaiaEDR3_phot_rp_mean_mag_error**: Error rp band magnitude (VEGA) from Gaia EDR3
51. **FUV**: Galex Far UV magnitude (AB magnitude)
52. **FUV_ERR**: Galex Far UV magnitude error (AB magnitude)
53. **NUV**: Galex Near UV magnitude (AB magnitude)
54. **NUV_ERR**: Galex Near UV magnitude error (AB magnitude)
55. **KiDS_u** : KiDS u-band magnitude (AB magnitude)
56. **KiDS_u_ERR**: KiDS u-band magnitude error (AB magnitude)
57. **KiDS_g**: KiDS g-band magnitude (AB magnitude)
58. **KiDS_g_ERR**: KiDS g-band magnitude error (AB magnitude)
59. **KiDS_r**: KiDS r-band magnitude (AB magnitude)
60. **KiDS_r_ERR**: KiDS r-band magnitude error (AB magnitude)
61. **KiDS_i**: KiDS i-band magnitude (AB magnitude)
62. **KiDS_i_ERR**: KiDS i-band magnitude error (AB magnitude)
63. **omegac_z**: OmegaCAM z-band magnitude (AB magnitude)
64. **omegac_z_ERR**: OmegaCAM z-band magnitude error (AB magnitude)
65. **HSC_g**: HSC g-band magnitude (AB magnitude)
66. **HSC_g_ERR**: HSC g-band magnitude error (AB magnitude)
67. **HSC_r**: HSC r-band magnitude (AB magnitude)
68. **HSC_r_ERR**: HSC r-band magnitude error (AB magnitude)
69. **HSC_r2**: HSC r2-band magnitude (AB magnitude)
70. **HSC_r2_ERR**: HSC r2-band magnitude error (AB magnitude)
71. **HSC_i**: HSC i-band magnitude (AB magnitude)
72. **HSC_i_ERR**: HSC i-band magnitude error (AB magnitude)
73. **HSC_i2**: HSC i2-band magnitude (AB magnitude)
74. **HSC_i2_ERR**: HSC i2-band magnitude error (AB magnitude)
75. **HSC_z**: HSC z-band magnitude (AB magnitude)
76. **HSC_z_ERR**: HSC z-band magnitude error (AB magnitude)
77. **HSC_Y**: HSC Y-band magnitude (AB magnitude)
78. **HSC_Y_ERR**: HSC Y-band magnitude error (AB magnitude)
79. **VIKING_z**: VISTA/VIKING z-band magnitude (AB magnitude)
80. **VIKING_z_ERR**: VISTA/VIKING z-band magnitude error (AB magnitude)
81. **VIKING_Y**: VISTA/VIKING Y-band magnitude (AB magnitude)
82. **VIKING_Y_ERR**: VISTA/VIKING Y-band magnitude error (AB magnitude)
83. **VIKING_J**: VISTA/VIKING J-band magnitude (AB magnitude)
84. **VIKING_J_ERR**: VISTA/VIKING J-band magnitude error (AB magnitude)
85. **VIKING_H**: VISTA/VIKING H-band magnitude (AB magnitude)
86. **VIKING_H_ERR**: VISTA/VIKING H-band magnitude error (AB magnitude)
87. **VIKING_Ks**: VISTA/VIKING Ks-band magnitude (AB magnitude)
88. **VIKING_Ks_ERR**: VISTA/VIKING Ks-band magnitude error (AB magnitude)
89. **W1**: LS8/Wise W1 magnitude (AB magnitude)
90. **W1_ERR**: LS8/Wise W1 magnitude error (AB magnitude)
91. **W2**: LS8/Wise W2 magnitude (AB magnitude)
92. **W2_ERR**: LS8/Wise W2 magnitude error (AB magnitude)
93. **W3**: LS8/Wise W3 magnitude (AB magnitude)
94. **W3_ERR**: LS8/Wise W3 magnitude error (AB magnitude)
95. **W4**: LS8/Wise W4 magnitude (AB magnitude)
96. **W4_ERR**: LS8/Wise W4 magnitude error (AB magnitude)
97. **LS8_g**: LS8 g-band magnitude (AB magnitude)
98. **LS8_g_ERR**: LS8 g-band magnitude error (AB magnitude)
99. **LS8_r**: LS8 r-band magnitude (AB magnitude)
100. **LS8_r_ERR**: LS8 r-band magnitude error (AB magnitude)
101. **LS8_z**: LS8 z-band magnitude (AB magnitude)
102. **LS8_z_ERR**: LS8 z-band magnitude error (AB magnitude)
103. **VHS_Y**: VISTA/VHS Y-band magnitude (AB magnitude)
104. **VHS_Y_ERR**: VISTA/VHS Y-band magnitude error (AB magnitude)
105. **VHS_H**: VISTA/VHS H-band magnitude (AB magnitude)
106. **VHS_H_ERR**: VISTA/VHS H-band magnitude error (AB magnitude)
107. **VHS_Ks**: VISTA/VHS Ks-band magnitude (AB magnitude)
108. **VHS_Ks_ERR**: VISTA/VHS Ks-band magnitude error (AB magnitude)
109. **HCS_g_diff**: Difference between psf and Kron magnitude in HSC g-band (AB magnitude)
110. **HCS_r_diff**: Difference between psf and Kron magnitude in HSC r-band (AB magnitude)
111. **HCS_i_diff**: Difference between psf and Kron magnitude in HSC i-band (AB magnitude)
112. **HCS_z_diff**: Difference between psf and Kron magnitude in HSC z-band (AB magnitude)

113. **HCS_opt_extended**: Extension in HSC griz bands. 1=extended; -99=data missing 0=other from Aihara et al 2018
114. TYPE:
115. **in_KiDS_flag**: Flag for KiDS coverage: 1: Source is in KiDS area; 0: otherwise
116. **in_HSC_flag**: Flag for HSC coverage: 1: Source is in HSC area as from Aihara et al 2018; 0: otherwise
117. **LS8_phot_flag**: Flag for LS8 photometry: 1: Source has simultaneously g,r,z,w1 photometry in LS8; 0: one or more bands are missing (boolean)
118. **SPECZ_RA**: Right Ascension (degrees) of the spectroscopic redshift entry in the original catalogue from which it was taken Right Ascension (degrees) of the spectroscopic redshift entry in the original catalogue from which it was taken
119. **SPECZ_Dec**: Declination (degrees) of the spectroscopic redshift entry in the original catalogue from which it was taken
120. **SPECZ_Redshift**: Spectroscopic redshift from original catalog
121. **SPECZ_NORMQ**: Normalised quality of spectroscopic redshift: 3=secure, 2=not secure, 1=unreliable redshift/bad spectrum, -1= Blazar candidate
122. **SPECZ_Origin**: Catalogue which provided this spectroscopic redshift
123. **SPECZ_Original_ID**: Identifier of this spectroscopic redshift entry in the original catalogue from which it was taken
124. **SPEC_Star_flag** : true when the CTP has a reliable redshift below 0.002 (boolean)
125. **SPEC_Gal_flag** : true when the CTP has a reliable redshift above 0.002
126. **CTP_class** :CTP Classification: SECURE EXTRAGALACTIC, LIKELY EXTRAGALACTIC, SECURE GALACTIC, LIKELY GALACTIC (see flowchart)
127. **PHZ_LEPHARE_zphot** : Photoz from Le PHARE, but set to 0 for GALACTIC sources
128. **PHZ_LEPHARE_zl68** : Le PHARE zphot min at 1 sigma
129. **PHZ_LEPHARE_zu68** : Le PHARE zphot max at 1 sigma
130. **PHZ_LEPHARE_zl90** : Le PHARE zphot min at 2 sigma
131. **PHZ_LEPHARE_zu90** : Le PHARE zphot max at 2 sigma
132. **PHZ_LEPHARE_zl99** : Le PHARE zphot min at 3 sigma
133. **PHZ_LEPHARE_zu99** : Le PHARE zphot max at 3 sigma
134. **PHZ_LEPHARE_chi** : Le PHARE chi2 value for best fitting galaxy/AGN template
135. **PHZ_LEPHARE_SED** : Le PHARE best template fitting the data
136. **PHZ_LEPHARE_extlaw** : Le PHARE Extinction Law applied to the template: Prevot (1) or none (0)
137. **PHZ_LEPHARE_ebv** : Le PHARE E(B-V) applied to the template
138. **PHZ_LEPHARE_pdz** : Le Phare probability distribution. Photoz more reliable when value is high
139. **PHZ_LEPHARE_nband** : Le Phare number of bands used for the computation of photoz
140. **PHZ_LEPHARE_zp2** : Le Phare second best photoz from LePhare, if existing
141. **PHZ_LEPHARE_chi2** : Le Phare chi2 value for second best fitting template, if existing
142. **PHZ_LEPHARE_SED2** : Le Phare second best template fitting the data, if existing
143. **PHZ_LEPHARE_pdz2** : Le Phare probability distribution for secondary solution, if existing
144. **PHZ_DNNz_zpho** : Photoz from DNNZ, but set to 0 for GALACTIC sources
145. **PHZ_DNNz_zl68** : DNNZ 1sigma min error on photoz
146. **PHZ_DNNz_zu68** : DNNZ 1 sigma max error on photoz
147. **PHZ_DNNz_zl95** : DNNZ 2 sigma min error on photoz
148. **PHZ_DNNz_zu95** : DNNZ 2 sigma max error on photoz
149. **CTP_REDSHIFT** : Final redshift: zspec (NORMQ=3) when available, else photo-z from Le PHARE; 0 for GALACTIC sources without spectroscopic redshift.
150. **CTP_REDSHIFT_GRADE** : In a range from 5 (spectroscopy) to 0 (unreliable photo-z) (see text)
151. **CLUSTER_CLASS**: In range from 5 to 1: 5=most likely a cluster; 1= not a cluster (see text)
152. **CTP_CLASSIFICATION** : same as CTP_class, but with numbers: 3: SECURE EXTRAGALACTIC; 2: LIKELY EXTRAGALACTIC; 1: SECURE GALACTIC; 0: LIKELY GALACTIC
153. **ID_MAIN**: [column present only in the Hard sample: The source ID in the main catalog for the sources in common.]

References

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, PASJ, 70, S4
Aihara, H., Armstrong, R., Bickerton, S., et al. 2018b, PASJ, 70, S8
Ananna, T. T., Salvato, M., LaMassa, S., et al. 2017, ApJ, 850, 66
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Baldry, I. K., Liske, J., Brown, M. J. I., et al. 2018, MNRAS, 474, 3875
Bianchi, L. 2014, Ap&SS, 354, 103
Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28
Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103
Brescia, M., Salvato, M., Cavuoti, S., et al. 2019, MNRAS, 489, 663
Brown, M. J. I., Duncan, K. J., Landt, H., et al. 2019, MNRAS, 489, 3351
Brusa, M., Comastri, A., Daddi, E., et al. 2005, A&A, 432, 69
Brusa, M., Zamorani, G., Comastri, A., et al. 2007, ApJS, 172, 353
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Civano, F., Elvis, M., Brusa, M., et al. 2012, ApJS, 201, 30
Comparat, J., Merloni, A., Dwelly, T., et al. 2020, A&A, 636, A97
Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 392, 19
de Vaucouleurs, G., de Vaucouleurs, A., Corwin, Herold G., J., et al. 1991, Third Reference Catalogue of Bright Galaxies
Dey, A., Schlegel, D. J., Lang, D., et al. 2019a, AJ, 157, 168
Dey, A., Schlegel, D. J., Lang, D., et al. 2019b, AJ, 157, 168
Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, arXiv e-prints, arXiv:1805.05489
Dotter, A. 2016, ApJS, 222, 8
Drinkwater, M. J., Byrne, Z. J., Blake, C., et al. 2018a, MNRAS, 474, 4151
Drinkwater, M. J., Byrne, Z. J., Blake, C., et al. 2018b, MNRAS, 474, 4151
Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, Astronomy and Geophysics, 50, 5.12
Dwelly, T., Salvato, M., Merloni, A., et al. 2017, MNRAS, 469, 1065
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37
Fotopoulou, S., Salvato, M., Hasinger, G., et al. 2012, ApJS, 198, 1
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2020, arXiv e-prints, arXiv:2012.01533
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Hickox, R. C. & Alexander, D. M. 2018, ARA&A, 56, 625
Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hoffleit, D. 1964, Catalogue of Bright Stars
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, ApJ, 796, 60
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683
Klein, M., Grandis, S., Mohr, J. J., et al. 2019, MNRAS, 488, 739
Klein, M., Mohr, J. J., Desai, S., et al. 2018, MNRAS, 474, 3324
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019a, A&A, 625, A2
Kuijken, K., Heymans, C., Dvornik, A., et al. 2019b, A&A, 625, A2
Lang, D. 2014, AJ, 147, 108

Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, Research in Astronomy and Astrophysics, 15, 1095
Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2010, ApJS, 187, 560
Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, ApJ, 326, 680
Marchesi, S., Civano, F., Elvis, M., et al. 2016, ApJ, 817, 34
McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, The Messenger, 154, 35
Meisner, A. M., Lang, D., Schlafly, E. F., & Schlegel, D. J. 2019, PASP, 131, 124504
Menzel, M.-L. et al. 2016, MNRAS, 457, 110
Merloni, A. 2016, Observing Supermassive Black Holes Across Cosmic Time: From Phenomenology to Physics, ed. F. Haardt, V. Gorini, U. Moschella, A. Treves, & M. Colpi, Vol. 905, 101
Merloni, A., Predehl, P., Becker, W., et al. 2012, ArXiv e-prints [arXiv:1209.3114]
Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1
Naylor, T., Broos, P. S., & Feigelson, E. D. 2013, ApJS, 209, 30
Noll, S., Mehlert, D., Appenzeller, I., et al. 2004, A&A, 418, 885
Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, A&A, 530, A122
Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, ApJS, 234, 34
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81
Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
Prevot, M. L., Lequeux, J., Maurice, E., Prevot, L., & Rocca-Volmerange, B. 1984, A&A, 132, 389
Rosen, S. R., Webb, N. A., Watson, M. G., et al. 2016, A&A, 590, A1
Ruiz, A., Corral, A., Mountrichas, G., & Georgantopoulos, I. 2018, A&A, 618, A52
Salvato, M., Buchner, J., Budavári, T., et al. 2018a, MNRAS, 473, 4937
Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, ApJ, 690, 1250
Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, ApJ, 742, 61
Salvato, M., Ilbert, O., & Hoyle, B. 2018b, Nature Astronomy
Shu, Y., Koposov, S. E., Evans, N. W., et al. 2019, MNRAS, 489, 4741
Simm, T., Saglia, R., Salvato, M., et al. 2015, A&A, 584, A106
Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32
Sutherland, W. & Saunders, W. 1992, MNRAS, 259, 413
Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
Taylor, M. B. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 666
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, ApJ, 475, 479
Willis, J. P., Oguri, M., Ramos-Ceja, M. E., et al. 2021, MNRAS, 503, 5624
Wolf, J., Nandra, K., Salvato, M., et al. 2021, arXiv e-prints, arXiv:2101.05585
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010b, AJ, 140, 1868
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010a, AJ, 140, 1868
York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579

---

[1] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr. 1, 85748 Garching, Germany
[2] Exzellenzcluster ORIGINS, Boltzmannstr. 2, D-85748 Garching, Germany
[3] Institute for Astronomy and Astrophysics, National Observatory of Athens, V. Paulou and I. Metaxa, 11532, Greece
[4] Dipartimento di Fisica e Astronomia "Augusto Righi", Universitá di Bologna, via Gobetti 93/2, 40129 Bologna, Italy
[5] INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, 40129 Bologna, Italy
[6] Department of Astronomy, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
[7] Academia Sinica Institute of Astronomy and Astrophysics, 11F of Astronomy-Mathematics Building, AS/NTU, No.1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
[8] Research Center for Space and Cosmic Evolution, Ehime University, 2-5 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan
[9] Leibniz-Institut für Astrophysik Potsdam (AIP). An der Sternwarte 16. 14482 Potsdam, Germany
[10] Universität Hamburg, Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany
[11] Institute for Advanced Research, Nagoya University Furocho, Chikusa-ku, Nagoya, 464-8602 Japan
[12] Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr 1, D-81679 Munich, Germany
[13] Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA True
[14] Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[15] Department of Physics, University of Connecticut, 2152 Hillside Road, Unit 3046, Storrs, CT 06269, USA
[16] Institute for Gravitation and the Cosmos, Pennsylvania State University, University Park, PA 16802, USA
[17] Yale Center for Astronomy and Astrophysics, Yale University, New Haven, CT, 06520, USA
[18] Instituto de Astronomıa, Universidad Nacional Autónoma de México A.P. 70-264, 04510, Mexico, D.F., México
[19] Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8578, Japan
[20] Astronomical Institute, Tohoku University, Aramaki, Aoba-ku, Sendai, Miyagi 980-8578, Japan
[21] Department of Astronomy and Astrophysics, The Pennsylvania State University,University Park, PA 16802
[22] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802 Dr. Karl Remeis-Observatory and Erlangen Centre for Astroparticle Physics, Sternwartstr. 7, 96049 Bamberg, Germany
[23] Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA
[24] Centro de Investigación en Astronomía, Universidad Bernardo O'Higgins, Avenida Viel 1497, Santiago, Chile
[25] Apache Point Observatory, P.O. Box 59, Sunspot, NM 88349
[26] Centro de Astronomía, Universidad de Antofagasta, Avenida Angamos 601, Antofagasta 1270300, Chile

Chapter 6

Application of the identified stellar X-ray sources

As an application of the identified sources, Freund et al. (2020) (Sect. 6.1) analyzed the X-ray properties of the Hyades cluster. With a distance of about 47 pc, the Hyades is the nearest well-populated cluster and have a medium age of $\sim$ 640 Myr. Since all members are coeval, open star clusters are ideally suited to investigate the activity-rotation-age relation. The X-ray properties of the Hyades were previously studied (e.g. Micela et al., 1988; Stern et al., 1995; Stelzer & Neuhäuser, 2001) before data from *XMM-Newton* and *Chandra* and a new processing of the RASS were available. Furthermore, *Gaia* strongly improved the membership identification and revealed the existence of tidal tails that have not been investigated in X-rays before. Due to the small distance of the Hyades cluster, their members cover a large area on the sky and pointings can only observe a small fraction of the members but they provide higher sensitivities. Therefore, all-sky surveys as RASS and eRASS are particularly important for the investigation of the X-ray properties of the Hyades.

We searched for X-ray detections of Hyades member from three different membership lists derived from *Gaia* DR2 data. Most X-ray detections are obtained by the RASS but also pointings from *ROSAT*, *Chandra*, and *XMM-Newton* are adopted. The RASS counterparts were taken from a preliminary version of the identifications presented in Freund et al., submitted (Sect. 4.1). The obtained results will be very helpful for later comparison with *eROSITA* data.

I combined the different membership lists and searched for X-ray detections and additional data myself, the *XMM-Newton* data were provided by P.C. Schneider. I also have the largest share in the interpretation of the results but, again, the coauthors provided important ideas to all processing steps and the analysis.

6.1 Publication: Updated X-ray view of the Hyades cluster

A&A 640, A66 (2020)
https://doi.org/10.1051/0004-6361/201937304

Astronomy & Astrophysics

Updated X-ray view of the Hyades cluster[⋆]

S. Freund, J. Robrade, P. C. Schneider, and J. H. M. M. Schmitt

Hamburger Sternwarte, Universität Hamburg, 21029 Hamburg, Germany
e-mail: `sebastian.freund@uni-hamburg.de`

Received 12 December 2019 / Accepted 4 June 2020

ABSTRACT

Aims. We revisit the X-ray properties of the main sequence Hyades members and the relation between X-ray emission and stellar rotation.

Methods. As an input catalog for Hyades members, we combined three recent Hyades membership lists derived from *Gaia* DR2 data that include the Hyades core and its tidal tails. We searched for X-ray detections of the main sequence Hyades members in the ROSAT all-sky survey, and pointings from ROSAT, the *Chandra* X-Ray Observatory, and *XMM-Newton*. Furthermore, we adopted rotation periods derived from *Kepler*'s K2 mission and other resources.

Results. We find an X-ray detection for 281 of 1066 bona fide main sequence Hyades members and provide statistical upper limits for the undetected sources. The majority of the X-ray detected stars are located in the Hyades core because of its generally smaller distance to the Sun. F- and G-type stars have the highest detection fraction (72%), while K- and M-type dwarfs have lower detection rates (22%). The X-ray luminosities of the detected members range from $\sim 2 \times 10^{27}$ erg s^{-1} for late M-type dwarfs to $\sim 2 \times 10^{30}$ erg s^{-1} for active binaries. The X-ray luminosity distribution functions formally differ for the members in the core and tidal tails, which is likely caused by a larger fraction of field stars in our Hyades tails sample. Compared to previous studies, our sample is slightly fainter in X-rays due to differences in the Hyades membership list used; furthermore, we extend the X-ray luminosity distribution to fainter luminosities. The X-ray activity of F- and G-type stars is well defined at $F_X/F_{\rm bol} \approx 10^{-5}$. The fractional X-ray luminosity and its spread increases to later spectral types reaching the saturation limit ($F_X/F_{\rm bol} \approx 10^{-3}$) for members later than spectral type M3. Confirming previous results, the X-ray flux varies by less than a factor of three between epochs for the 104 Hyades members with multiple epoch data, significantly less than expected from solar-like activity cycles. Rotation periods are found for 204 Hyades members, with about half of them being detected in X-rays. The activity-rotation relation derived for the coeval Hyades members has properties very similar to those obtained by other authors investigating stars of different ages.

Key words. open clusters and associations: individual: Hyades – X-rays: stars – stars: activity – stars: rotation – stars: coronae – stars: late-type

1. Introduction

Open star clusters provide ideal laboratories to study the activity-rotation-age relation of late-type stars because all their members are coeval. As the nearest well-populated open cluster, the Hyades with a distance of only ~47 pc from the Sun and an age of about 640 Myr (Lodieu et al. 2019) is of unique importance. The second *Gaia* data release (*Gaia* DR2 Gaia Collaboration 2018b, 2016), containing highly accurate parallaxes and proper motions for 1.3 billion sources and radial velocities for about 7 million sources, also significantly improves the membership identification of the Hyades. Adopting these data, Lodieu et al. (2019) present a revised census of the members of the Hyades core. According to Röser et al. (2011), Hyades members within a distance of 9 pc are gravitationally bound, but for stars at larger distances, the Galaxy exerts tidal forces leading to the creation of tidal stellar tails, which have been discovered independently by Meingast & Alves (2019) and Röser et al. (2019) using data from *Gaia* DR2.

The investigation of rotation periods for Hyades members has a long history and the Hyades are among the first open clusters for which photometric rotation periods were measured for low-mass stars (Radick et al. 1987, 1995). Since then, the number and quality of available rotation periods has increased substantially especially due to *Kepler*'s K2 mission (Howell et al. 2014). With its ~100 deg^2 field of view, *Kepler* targeted two slightly different parts of the Hyades in K2 Campaign 4 and 13 for about 75 days each. Douglas et al. (2016, 2019) use these K2 data to derive new rotation periods for Hyades members (for more details see Sect. 2.2).

X-ray emission from the Hyades members has been systematically investigated ever since the *Einstein* Observatory era (Stern et al. 1981; Micela et al. 1988). As a Hyades input catalog, Micela et al. (1988) combined 323 certain or probable Hyades members from different optical catalogs, arguing that their catalog is complete down to the ninth magnitude. Of Hyades members covered by a pointing of the *Einstein* Observatory, 66 of 121 are detected, and depending on the individual exposure time and the off-axis angle of the considered Hyades, Micela et al. (1988) reach detection limits of $\sim 2{-}7 \times 10^{28}$ erg s^{-1} at the assumed Hyades distance of 45 pc. The first complete X-ray survey of the Hyades cluster region was performed by Stern et al. (1995) using data from the ROSAT all-sky survey (RASS). They discuss the X-ray properties of 440 optically selected Hyades members, including fainter sources than the input catalog of Micela et al. (1988). With an estimated detection limit of $\sim 1.5{-}3 \times 10^{28}$ erg s^{-1} at 45 pc, Stern et al. (1995) detect 187 of their Hyades members as X-ray sources and derive upper limits for the undetected sources. ROSAT Position Sensitive Proportional Counter (PSPC) pointed observation of the

---

[⋆] Data are only available at the CDS via anonymous ftp to `cdsarc.u-strasbg.fr` (`130.79.128.5`) or via `http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/640/A66`

Taurus-Auriga-Perseus region including the Hyades are analyzed by Stelzer & Neuhäuser (2001). Now, new data reductions for the RASS (Boller et al. 2016) and ROSAT PSPC pointed observations (ROSAT 2000) are available, and furthermore observations with *XMM-Newton* and the *Chandra* X-Ray Observatory (Weisskopf et al. 2002) provide new Hyades detections with unprecedented sensitivity and accuracy allowing us to detect sources with later spectral types and lower activity.

In this paper, we revisit the X-ray properties and the activity-rotation relation of the Hyades cluster applying current membership lists including the tidal tails, rotation periods, and significantly improved X-ray detection lists. We discuss the data acquisition in Sect. 2 describing the membership list of the Hyades in Sect. 2.1, the rotation periods in Sect. 2.2, and the X-ray observations from the different instruments in Sect. 2.3. In Sect. 3 we present our results discussing the X-ray detection rates in Sect. 3.1, the distribution of the X-ray luminosity and activity in Sect. 3.3, the variability of the Hyades members in Sect. 3.4, and the activity-rotation relation in Sect. 3.5. Finally, we draw our conclusions in Sect. 4.

2. Data acquisition

We acquired our data from different resources and created our Hyades membership list by combining recent publications from Lodieu et al. (2019), Meingast & Alves (2019), and Röser et al. (2019). To obtain the rotational properties of the Hyades members, we adopted the rotation periods from Douglas et al. (2019), who combine data from multiple studies and added further periods from Lanzafame et al. (2018). Furthermore, we cross-matched our Hyades membership list with X-ray detections from the RASS, and pointings from ROSAT, the *Chandra* X-Ray Observatory, and *XMM-Newton*.

2.1. Membership

We obtained our new Hyades member sample from three recent publications that identify Hyades members based on data from *Gaia* DR2. Lodieu et al. (2019) studied the central region of the Hyades and associate 710 *Gaia* sources with the Hyades, located within 30 pc to the cluster center; they estimate the contamination of this sample to be around 5−10%.

The tidal tails of the Hyades were independently discovered by Meingast & Alves (2019) and Röser et al. (2019). Both publications started from the same data, which were the *Gaia* DR2 sources within a 200 pc sphere around the Sun, but different selection criteria were adopted to extract the Hyades core and tidal tail members. The most important difference is that Röser et al. (2019) identified the Hyades members by applying the convergent-point method solely based on tangential velocities, while Meingast & Alves (2019) relied on 3D space velocities, excluding all sources without a radial velocity measured in *Gaia* DR2. However, radial velocities are only available for *Gaia* DR2 sources with G magnitudes between 4 and 13 mag and effective temperatures in the range of 3550–6900 K. Röser et al. (2019) subdivide their sample in the core containing sources within 18 pc around the cluster center, and they further differentiate between the leading and trailing tail selected by eye.

For the following analysis, we adopted the 979 sources from Röser et al. (2019) classified as core or tail sources and not flagged as possible interlopers. According to Röser et al. (2019), ∼1.4% and ∼13% of the sources in the core and in the tails are spuriously associated to the Hyades. Furthermore, we obtained 238 Hyades members from Meingast & Alves (2019), who do not comment on the reliability of their sample.

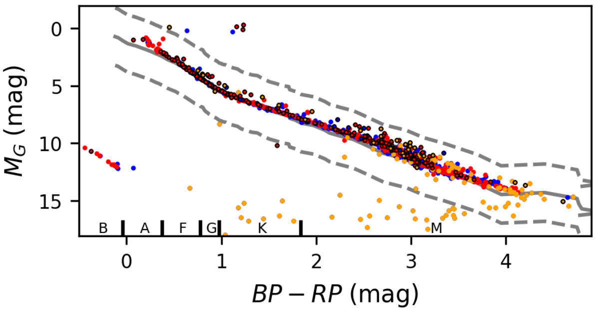


Fig. 1. Hertzsprung-Russell diagram of the Hyades. The red and blue sources indicate Hyades members in the core and in the tails, respectively. The orange sources are only associated to the Hyades by Lodieu et al. (2019), sources outlined in black are detected X-rays. The theoretical main sequence adopted from Worthey & Lee (2011) is shown by the black solid line, while the dashed lines indicate the values 2.5 mag brighter and fainter than the main sequence.

Table 1. Catalogs of the Hyades.

Catalog	*N*	Catalog	*N*
L	108	R	360
M	17	LR	361
LM	6	RM	46
LRM	168	Sum	1066

References. L: Lodieu et al. (2019), R: Röser et al. (2019), M: Meingast & Alves (2019).

Combining the three Hyades catalogs, we obtained 1142 unique bona fide members. The Hertzsprung-Russell diagram (HRD) of the Hyades members selected in this way (Fig. 1) shows that most of the sources are located on the main sequence. The four Hyades giants are located well above the main sequence (and discussed in Schröder et al. 2020) and the white dwarfs are found at $M_G < 10$ mag and $BP - RP < 0$ mag in the HRD. Furthermore, there are many sources that are clearly fainter than expected for main sequence stars but have red colors. We inspected these sources and found that they are only associated to the Hyades by Lodieu et al. (2019) and are excluded by Röser et al. (2019) because they do not meet the quality criteria of a small "unit weight error" and "flux excess ratio" proposed by Lindegren et al. (2018) and Gaia Collaboration (2018a). Therefore, the positions of these sources in Fig. 1 are probably wrong. In the following analysis we therefore concentrate on main sequence stars, excluding all sources that are 2.5 mag brighter or fainter than expected for dwarfs, thus obtaining our sample of 1066 bona fide Hyades members. This brightness cut excludes most of the rather peculiar red sources below the main sequence (cf., Fig. 1). We expect that the remaining sources are highly probable Hyades members thanks to their location on or very close to the Hyades main sequence, although some do not meet the formal "unit weight error" and "flux excess ratio" criteria.

In Table 1 we list the catalogs providing the Hyades members. To distinguish between core and tidal tails, we applied the condition of Röser et al. (2019) and classify Hyades member within 18 pc of the cluster center as core sources and

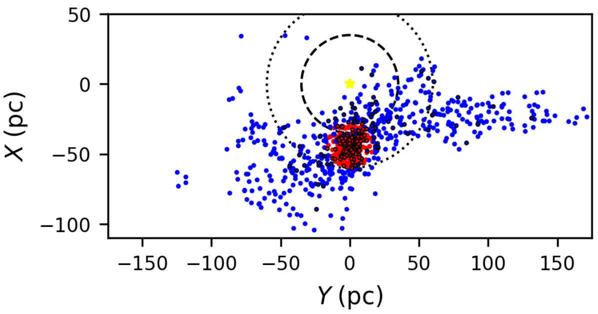


Fig. 2. Hyades members in Galactic Cartesian XY coordinates. The red and blue sources are associated with the cluster's core and tidal tails, respectively. Sources outlined in black are detected in X-rays. The star marker indicates the position of the sun and the dashed and dotted lines show 35 pc and 60 pc radius around the sun, respectively.

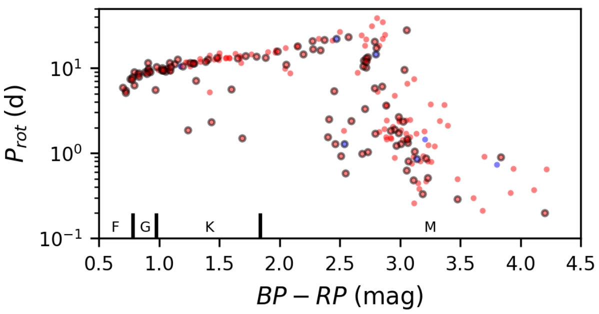


Fig. 3. Rotation periods of the Hyades as a function of the $BP - RP$ color. The red and blue sources are associated to the core and the tidal tails, respectively. Sources outlined in black are detected in X-rays.

members with larger distances as tidal tail sources, adopting the position of the cluster center from Lodieu et al. (2019). Thus, 550 Hyades are associated with the core and 516 with the tidal tails. In Fig. 2 we show the spatial distribution of the Hyades members discussed in this paper in Galactic Cartesian coordinates, where the X and Y axes are directed to the Galactic center and in the direction of the Galactic rotation, respectively. Most core sources have similar distances, some tail members are in the immediate solar vicinity, most tail members, however, have much larger distances than the Hyades core.

2.2. Rotation periods

Douglas et al. (2019) provide rotation periods for Hyades members derived from K2 Campaign 13 light curves. Additionally, they adopt rotation periods from Douglas et al. (2016) estimated from K2 Campaign 4 data as well as periods from Radick et al. (1987, 1995), Prosser et al. (1995), Hartman et al. (2011), and Delorme et al. (2011), and from data of the All Sky Automated Survey (ASAS, Pojmanski 2002). We searched Table 3 of Douglas et al. (2019) for matches with our membership list and adopted 191 proposed rotation periods. For 13 Hyades members not listed by Douglas et al. (2019), we found a rotation period in Lanzafame et al. (2018) who analyze photometric time series from *Gaia* data (released within DR2). Thus in total we have rotation periods for 204 Hyades members. Since Douglas et al. (2019) only investigate the Hyades core, the great majority of the stars with periods are located in the core and only seven of the tail members have known periods.

In Fig. 3 we show the rotation periods as a function of the $BP-RP$ color. From F-type to early M-type stars, the sources follow the slow-rotator sequence with a steady increase in the rotation period to later spectral types. There are a few outliers with smaller periods; these sources are probably binaries. The rotation period abruptly drops at $BP - RP \approx 2.5$ mag near the boundary to fully convective stars (compare to Douglas et al. (2019)). The X-ray detected Hyades are marked with black symbols in Fig. 3 and we discuss their properties in Sect. 3.5.

2.3. X-ray observations

We adopted the X-ray properties by crossmatching the Hyades members with the second ROSAT all-sky survey (2RXS) source catalog (Boller et al. 2016), the second ROSAT PSPC catalog (release 2.1.0 ROSAT 2000) (2RXP), the ROSAT High Resolution Imager (HRI) pointed observations (3rd release ROSAT Scientific Team 2000) (1RXH), the *Chandra* Source Catalog (specifically CSC release 2.0 Evans et al. 2010), and pointed *XMM-Newton* observations that contain a Hyades member. For sources with multiple detections, we adopted the X-ray data derived from *XMM-Newton* data or from the CSC if available, and otherwise the ROSAT observation with the longest exposure time. To make the measurements of the different instruments comparable, we converted all X-ray fluxes into the *XMM-Newton* band (0.2−12 keV), adopting an Astrophysical Plasma Emission Code (APEC) thermal plasma model with a temperature of $\log(T) = 6.5$ and solar metallicity; we also estimated upper limits for the undetected sources using this plasma emission model. We note that we derived these upper limits only in a statistical way, not individually using the actual photon counts; for any given star, an actual upper limit might be higher, if, for example, a nearby source is present, or lower if the background is exceptionally low. Our upper limits therefore characterize the properties of our Hyades sample. We did compare our upper limits with those given by Stern et al. (1995) and found in most cases our values to be rather conservative. Hence, we are confident that our upper limits are reliable for most Hyades members.

2.3.1. ROSAT all-sky survey

Between August 1990 and January 1991 the RASS was performed (Truemper 1982). Boller et al. (2016) re-processed the data from this survey and created the 2RXS catalog. We adopted the 2RXS identifications of the Hyades members from Freund et al. (in prep.), considering only those members that are the most likely counterpart to the 2RXS source and have an individual matching probability $p(H_{1j}) > 50\%$.

In contrast to pointed observations, the RASS covers the entire region of the Hyades, and hence, upper limits of the X-ray luminosity can be derived for all undetected Hyades members. To estimate the minimal number of counts that are detectable in the RASS, we inspected the number of detected counts as a function of exposure time for all 2RXS detections as shown in Fig. 4. We fitted a lower envelope to the distribution using a quadratic polynomial so that 95 % of the sources are located above the envelope. Then, we applied the RASS exposure time at the position of every undetected Hyades member and the lower envelope to estimate the minimal number of counts that would have resulted in an X-ray detection, and thus, the upper limit of the count rate and the X-ray luminosity.

2.3.2. ROSAT pointings

After the all-sky survey, ROSAT performed pointed observations until the end of the mission in February 1998. For these observations two different instruments were available: the Position Sensitive Proportional Counter (PSPC) and the High Resolution Imager (HRI). The results are available in the 2RXP (ROSAT 2000) and 1RXH (ROSAT Scientific Team 2000) catalogs, respectively.

We crossmatched the Hyades member list with the 2RXP and 1RXH catalogs, applying a matching radius of 3σ of the stated X-ray positional uncertainty, excluding all 2RXP and 1RXH sources with no exposure time and 1RXH sources flagged as non-unique. No positional and count rate errors are given for 21% of the 2RXP sources detected at an off-axis angle larger than 20 arcmin to the center of the field of view. For these sources, we estimated the positional uncertainty considering the width of the point spread function (PSF) of the Gaussian intrinsic resolution and the mirror blur[1] through

$$\sigma = \frac{\sqrt{108.7E^{-0.888} + 1.121E^{6} + 0.219\theta^{2.848}}}{\sqrt{\mathrm{cts}}}\ \mathrm{arcsec}, \qquad (1)$$

where cts is the number of source counts, θ is the off-axis angle in degrees, and for the source energy, we adopted a constant value of $E = 0.2$ keV. For the 2RXP and 1RXH sources with a formal statistical uncertainty <5 arcsec, we applied a value of 5 arcsec to account for systematic errors. Some Hyades members are counterparts to multiple 2RXP or 1RXH sources. Here, we calculated the weighted mean of the count rates from all 2RXP or 1RXH sources because they are likely detections of the same X-ray source.

For the Hyades members covered by a 2RXP or 1RXH pointing and not detected in any of the X-ray catalogs, we estimated upper limits. We adopted a circular field of view with a diameter of 2° and 38 arcmin for the 2RXP and 1RXH pointings, respectively, neglecting the ribs of the PSPC and the square field of view of the HRI. We estimated the minimal number of detectable counts as a function of the exposure time in a similar way to that described in Sect. 2.3.1. However, since the width of the PSF strongly increases with the off-axis angle, we applied different lower envelopes to 2RXP sources depending on their off-axis angle. We derive 2RXP upper limits only for Hyades members within 50 arcmin to the center of a 2RXP pointing because the number and quality of sources detected at larger angular separations is too low to fit a lower envelope and derive meaningful upper limits.

2.3.3. *Chandra* pointings

The *Chandra* X-Ray Observatory has been performing pointed observations since its launch in 1999 (Weisskopf et al. 2002). Two different instruments, the Advanced CCD Imaging Spectrometer (ACIS) and the High Resolution Camera (HRC), are available for the observations. The CSC 2.0 provides information on about 370 000 unique sources detected in more than 10 000 *Chandra* observations with the ACIS and HRC instruments; this information was made publicly available by the end of 2014. Pointings covering the same position were stacked to determine unique sources. In the region of the Hyades, *Chandra* pointings cover a region of about 0.70 deg^2 in the core and 0.47 deg^2 in the tails. We searched in the CSC for counterparts and upper limits to our Hyades members, adopted a matching radius of 3σ of the *Chandra* positional uncertainties, and applied a minimal uncertainty of 1 arcsec to account for systematic errors. We did not search for individual observations. The CSC provides X-ray fluxes estimated from an APEC model and we adopted these values after converting the flux of the broad bands of the ACIS and HRC instruments to the XMM band.

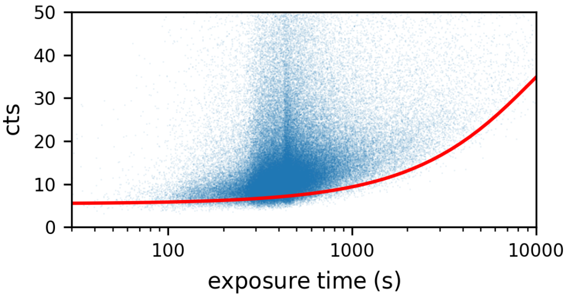


Fig. 4. Number of counts as a function of the exposure time for all 2RXS sources. The solid red curve represents the lower envelope (95% of the sources lie above the line).

2.3.4. *XMM-Newton* pointings

We inspected all *XMM-Newton* observations with a Hyades member that were within 10 arcmin of the nominal aim point of the X-ray observation covering a region of about 6.7 and 1.2,deg^2 of the Hyades core and tails, respectively. If the source fell onto one of the three X-ray detectors, we extracted the source and background counts within 10 and 20 arcsec from the proper-motion-corrected position on the detector(s). If the number of counts in any of the two considered radii was above the 99% interval for random background fluctuations, we considered the source as potentially detected. In that case, we considered all *Gaia* DR2 sources that fell within 20 arcsec of the Hyades member under consideration and checked if any of these potential alternative sources provided a better match with the X-ray photon distribution. For example, if the centroid matched better with any of the alternative sources, we manually checked whether the Hyades member was indeed the best matching counterpart (see Schneider et al, in prep. for details).

Hyades members that are detected by *XMM-Newton* were then checked against source variability during the observation and the source count rate was converted into flux using the encircled energy fraction, the effective exposure at the detector position, and assuming a plasma temperature of $\log(T) = 6.5$ as used for the ROSAT conversion factor.

3. X-ray and rotational properties of the Hyades

In Table 2 we provide the number of Hyades detections, unique sources, and data adopted for the following analysis for the individual X-ray catalogs as well as the number of upper limits and number of lowest upper limits from all X-ray catalogs. In total, 281 Hyades members are detected as X-ray sources. In the following, we discuss the X-ray detection rate and the X-ray and rotational properties of the Hyades members in detail.

3.1. X-ray detection rates

Most of the detections (212) and upper limits from pointed observations (77) are located in the core of the Hyades, while only 69 and 30 detections and upper limits are found in the tidal

---

[1] `https://heasarc.gsfc.nasa.gov/docs/journal/rosat_off-axis_psf4.html`

Table 2. Catalogs of the X-ray data for the Hyades.

Catalog	Detections	Unique sources	Adopted data	Upper limits	Lowest upper limits
2RXS	191	191	117	869	678
2RXP	155	103	85	156	91
XMM	63	58	58	4	4
1RXH	44	39	19	22	10
CSC	-	7	2	3	2

Table 3. X-ray detections of the Hyades with different distances to the Sun.

	$d_\odot < 35$ pc			$35 < d_\odot < 60$ pc			$d_\odot > 60$ pc		
SpT	All	Detected	Fraction [%]	All	Detected	Fraction [%]	All	Detected	Fraction [%]
A	1	0	0	3	0	0	0	0	–
F	2	1	50	37	34	92	15	6	40
G	1	0	0	28	25	89	14	6	43
K	14	7	50	96	38	40	46	3	7
M	47	13	28	505	131	26	257	17	7

tails. The reason for the smaller number of detections in the tails is the generally larger distance of the tail sources. As evident in Fig. 2, most of the detected tail sources are rather close to the Sun and only a very few Hyades members are detected at larger distances. Furthermore, the central part of the Hyades cluster is much better covered by pointed X-ray observations. To further investigate the influence of the source distances on the detection fractions, we sorted the Hyades members into three groups according to their distance. The first group contains 65 sources closer than 35 pc to the Sun, with 5 sources being associated to the core and the remaining 60 to the tidal tails. The second group contains stars with distances between 35 and 60 pc and covers the center of the Hyades; hence 526 of the so-selected 669 sources are classified as core sources, while 143 objects belong to the tidal tails. The third group of stars with distances larger than 60 pc contains more sources in the tails (313), while only 19 are classified as core sources.

In Table 3 we provide the detection fractions of the different samples. As expected, the sample with the largest distance has the lowest detection fraction and the closest group has the highest detection fraction. Furthermore, the detection fraction depends on the spectral type (as adopted from the $BP-RP$ color). While the largest detection fractions are obtained for F- and G-type stars, the detection fraction drops for K- and M-type members due to the X-ray luminosity distributions of the different spectral types (cf., Sect. 3.3). We find no X-ray emitting A-type Hyades members, confirming the fact that A-type stars are in general not strong X-ray sources (Schmitt 1997).

We find 291 of the 440 sources analyzed by Stern et al. (1995) in our sample of Hyades dwarfs (a further 34 members in Stern et al. (1995) have an identification in our Hyades catalog, but are not located on the main sequence). Stern et al. (1995) identify 111 sources with a RASS counterpart and we add a further 56 X-ray identifications of these sources. On the other hand, we do not find a reliable X-ray identification for 10 Hyades members reported by Stern et al. (1995) for various reasons; in some cases the angular separation between the Hyades member and the RASS counterpart is quite large and therefore we deem the identification as unreliable, or there is a better counterpart to the RASS source, either a binary companion or a background source unrelated to the Hyades cluster. A few RASS sources found by Stern et al. (1995) are not part of the 2RXS catalog that we use for our study.

3.2. Spurious identifications

Freund et al. (in prep.) provide matching probabilities for all 2RXS identifications of the Hyades members. Thus, we expect 13 of the 191 2RXS identifications to be spurious associations, however, the estimation of the matching probability does not consider that we know a priori Hyades members to be likely identifications. Hence, we expect the estimated number to be an upper bound.

To estimate the number of random 2RXP, 1RXH, XMM, and CSC associations, we randomly shifted all Hyades members several times between 3 and 10 arcmin and by a random angle, and performed the same matching procedure for the shifted as for the real Hyades members. On average, we obtained 2.2 2RXP and 0.2 1RXH identifications to the shifted Hyades members and we expect a similar number of random associations for the real Hyades members. The probability of a chance alignment for XMM and CSC counterparts is thus very low (<1 spurious counterparts in our sample) according to our tests.

3.3. X-ray activity and Hertzsprung-Russell diagram

In Figs. 5 and 6 we show the fractional X-ray fluxes F_X/F_{bol}[2] and X-ray luminosities L_X of the main sequence Hyades members as a function of the $BP-RP$ color. The activity distribution of the F- and G-type stars is quite well defined, but the spread in fractional X-ray luminosity increases for later spectral types, reaching a maximum for K- and early M-type sources. Nevertheless, the X-ray activity (as measured by F_X/F_{bol}) continuously increases toward later spectral types, from $\log(F_X/F_{\mathrm{bol}}) \approx -5.7$

---

[2] For the bolometric correction and the conversion between the $BP-RP$ and other photometric colors, throughout this page we adopted the values given in a table based on Pecaut & Mamajek (2013) available at `http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt` (Version 2019.3.22).

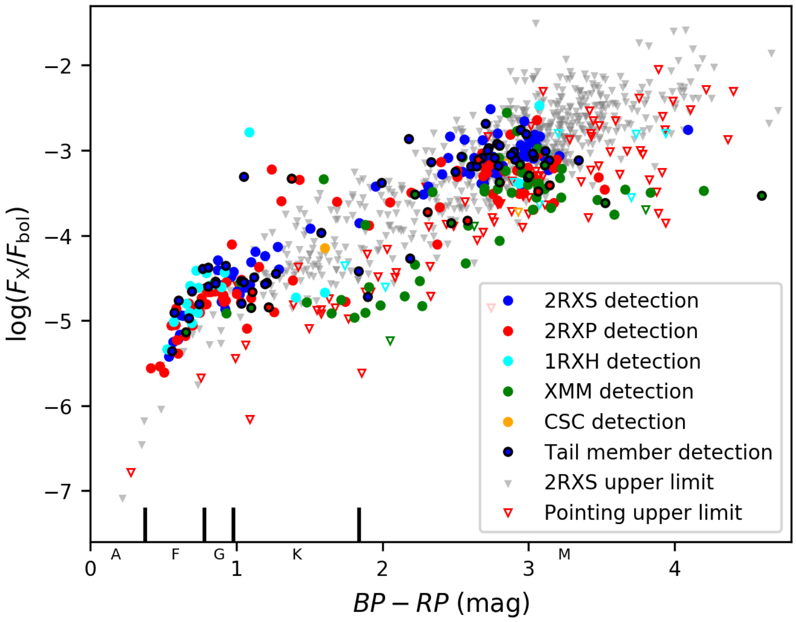


Fig. 5. X-ray to bolometric flux distribution of the Hyades as a function of the $BP-RP$ color. Blue, red, cyan, green, and orange dots indicate the values adopted from 2RXS, 2RXP, 1RXH, XMM, and CSC, respectively, while the triangles show the upper limits from 2RXP, 1RXH, and XMM pointings. The upper limits from 2RXS are shown as gray triangles in the background. In contrast to the detections in the core, the Hyades members in the tails are outlined in black. At the bottom we show the ranges of the spectral types as a guidance.

for early F-type stars to the saturation limit at $\log(F_X/F_{\rm bol}) \approx -3$ for sources later than approximately spectral type M3 (Vilhu 1984; Agrawal et al. 1986; Fleming et al. 1988; Pallavicini et al. 1990). Some sources at late spectral types are detected above the saturation limit, which we tentatively attribute to flares.

In Fig. 6 some early K-type sources with X-ray luminosities higher than $L_X = 5 \times 10^{29}$ erg s^{-1} stand out particularly. These sources are X-ray bright, active binaries like RS CVn systems; close binaries are known to be very active in X-rays because they maintain their high rotation periods due to tidal interaction (Walter et al. 1978; Dempsey et al. 1993, 1997). On the other hand, Fig. 6 includes three upper limits with $L_X \lesssim 10^{27}$ erg s^{-1}, which would be highly surprising given the age of the Hyades members. We inspected these sources in detail and found that one data point (*Gaia* DR2 3312602348628348032 at $BP-RP = 1.1$ mag) is probably erroneously an upper limit as a likely counterpart is just outside the positional error region (reasonable given the number of X-ray detections). The two other low upper limits (at $BP-RP = 1.9$ and 2.7 mag) are likely indicative of the fact that these sources are not genuine Hyades members (specifically *Gaia* DR2 3189577958236917632 and 45770783575075968), which we regard as compatible with the estimated contamination fraction in the membership list of several percent.

After excluding these three upper limits, the faintest Hyades members detected with *XMM-Newton* have X-ray luminosities that are similar to the lowest upper limits. Hence, we conclude that all main sequence Hyades stars with a convection zone can be detected in X-rays at sensitivities reachable with *XMM-Newton*, with the only exception possibly being the latest M-type stars.

In Fig. 7 we compare the X-ray luminosity distributions of the core and tail members of the Hyades for different spectral types and show the distributions resulting from the data by Stern et al. (1995) as comparison; we used the Kaplan-Meier estimator to include upper limits, which is of particular importance because the detection limits of the tail sources are higher than in the core due to their generally larger distances. As visible in Fig. 7, the earlier type stars have a steeper X-ray luminosity distribution, and we note the different scaling of the x-axis in the upper panels. For most spectral types, the core members are slightly brighter than the members in the Hyades tails. According to a logrank test, the X-ray luminosity distributions of the core and tail members differ with a high significance. However, this does not necessarily mean that the X-ray properties of the Hyades core and tail members differ intrinsically, instead the difference is probably caused by a larger fraction of field stars spuriously associated with the Hyades tidal tails as discussed by Röser et al. (2019). Even our core members are fainter than the sample presented by Stern et al. (1995). The deviations are mainly caused by the different Hyades membership lists; for example, four of the five brightest F-type X-ray emitters in Stern et al. (1995) are not associated to the Hyades according to the *Gaia* DR2 data that we use. However, due to the detections from pointed observations especially with *XMM-Newton*, we can extend the X-ray luminosity distribution to fainter luminosities, which is particularly important for the K- and M-type Hyades members.

3.4. Variability

X-ray luminosities from different epochs are available for 104 Hyades members detected by different instruments or by the same instrument in different pointings. In the bottom panel of Fig. 6 and in Fig. 8, we compare the X-ray luminosities of the various detections. We do not include upper limits because our upper limits are suited to describe the properties of samples but they might be unreliable for individual sources. Although the observations are up to 25 years apart, in most cases the X-ray luminosities measured for each source do not differ by more than a factor of three. Hence we conclude that the Hyades members are not strongly affected by activity cycles compared to the Sun, whose X-ray luminosity differs by more than one order of magnitude between solar minimum and maximum (Peres et al. 2000) in a comparable energy band.

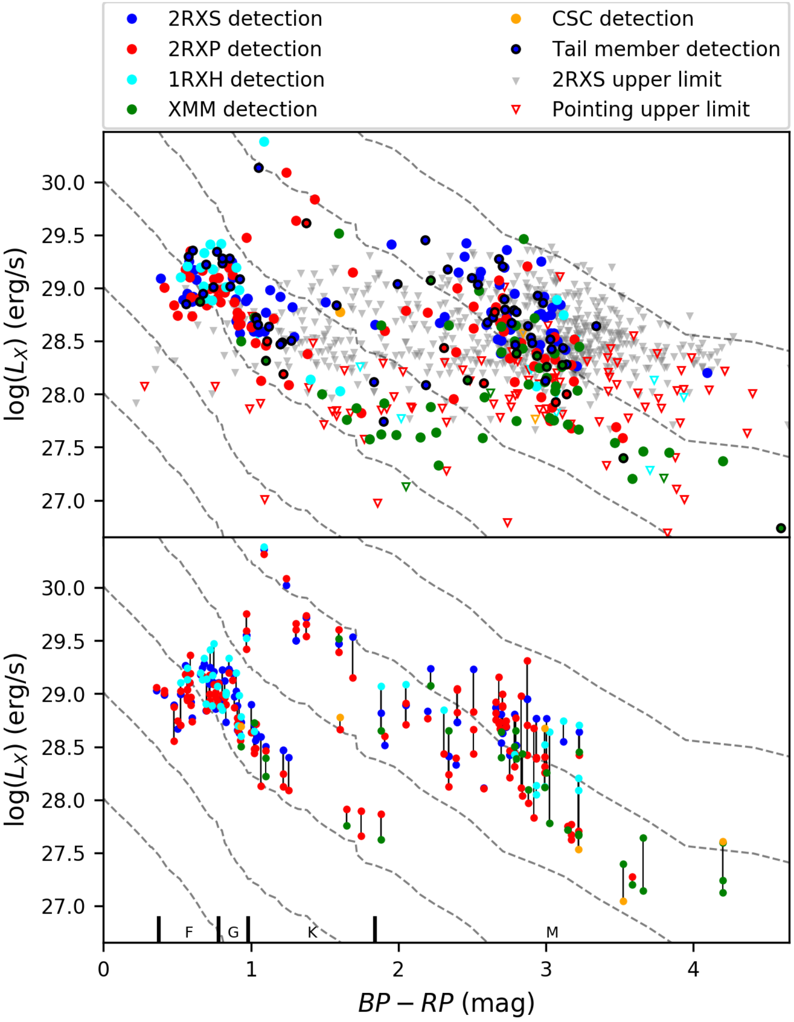
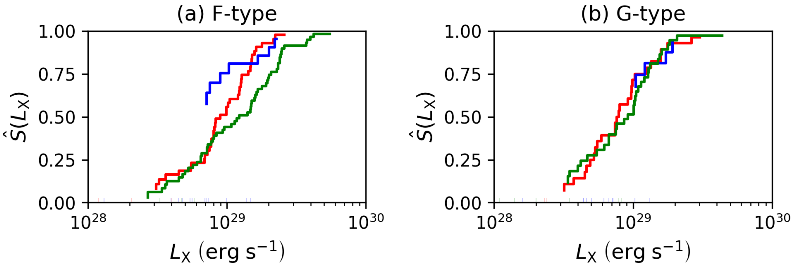


Fig. 6. X-ray luminosities of the Hyades as a function of the $BP-RP$ color. The color coding is the same as in Fig. 5. From top to bottom, the dashed lines indicate activity levels of $L_X/L_{\mathrm{bol}} = -2$, -3, -4, -5, -6, and -7 for sources located on the main sequence. The upper panel shows the best X-ray luminosity for the individual Hyades members, while the bottom panel compares the X-ray luminosities measured by different instruments or pointings for multiple detected sources. Detections of the same source are connected by a string.


(a) F-type (b) G-type

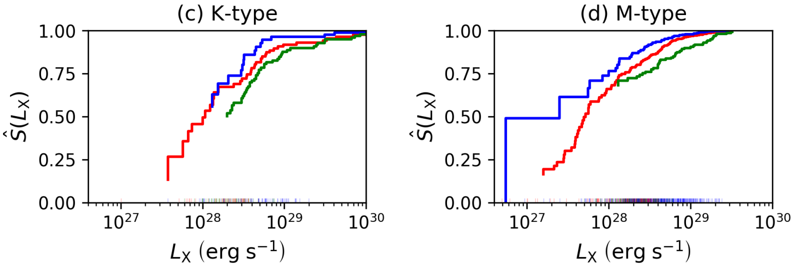

(c) K-type (d) M-type

Fig. 7. Kaplan-Meier estimator of the X-ray luminosities for the Hyades members at different spectral types. The red, blue, and green lines show the survival functions for the core and tail members and for the sample from Stern et al. (1995), respectively. The colored bars at the x-axis indicate the upper limits

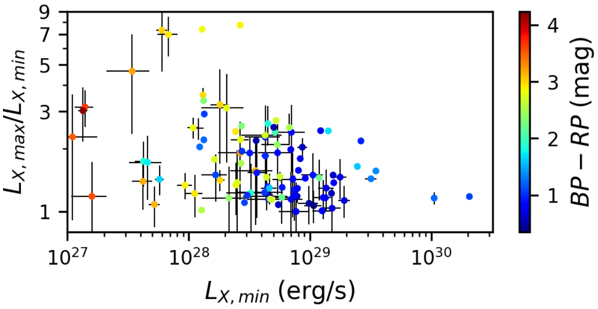


Fig. 8. Fraction of the maximal to minimal measured X-ray luminosity for multiple detected Hyades members. The color scales with the $BP - RP$ color. No error is given for 2RXP sources detected at an off-axis angle larger than 20 arcmin to the center of the pointing.

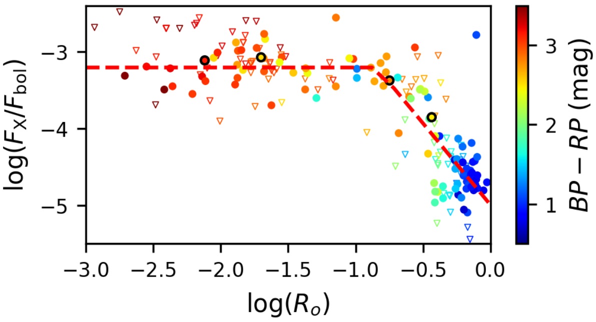


Fig. 9. X-ray activity as a function of the Rossby number for Hyades members with known rotation period. The dots indicate sources detected in X-rays, while the triangles show upper limits. In contrast to the sources in the core, the Hyades members in the tails are outlined in black. The color scales with the $BP - RP$ color. The red dashed line indicates the best fit.

Hyades members that show larger differences between epochs are particularly red and, therefore, low-mass stars. These stars are generally more active and therefore show more frequent flaring. Such stochastic flares can increase the X-ray flux by a factor of more than 100 (e.g., Stelzer et al. 2006). Hence, the larger differences in the X-ray fluxes derived at two epochs are likely caused by different degrees of flaring activity.

This agrees well with the results reported by Stern et al. (1995), who find most of the Hyades members to vary by not more than a factor of two and attributed most of the larger variations to flares. Similarly, Micela et al. (1996) find a variability by a factor of two for only about 15% of the much younger Pleiades members; thus X-ray flaring does not corrupt these cross-mission comparisons.

3.5. Rotation properties

Out of 204 Hyades members with measured rotation periods, 103 are detected in X-rays. As shown in Fig. 3, the detection fraction generally decreases with increasing rotation period. Hence, most of the potential binaries located below the slow-rotator sequence and many of the earlier fully convective stars are detected in X-rays. However, despite their short rotation periods, the detection fraction decreases for the latest spectral types because of their low bolometric luminosities.

For the Hyades members with known rotation periods, we estimated the Rossby number by adopting empirical convective turnover times provided by Eq. (5) in Wright et al. (2018), which is an improved version of Eq. (10) from Wright et al. (2011). However, for four Hyades members with known rotation periods, the color is outside the range in which the correlation from Wright et al. (2018) is valid. In Fig. 9 the X-ray activity as measured by F_X/F_{bol} is plotted as a function of the Rossby number. Most sources with $\log(R_o) \lesssim -1.0$ are saturated; these sources are generally the later type stars. The activity linearly decreases for larger Rossby numbers and stars of earlier spectral types also have larger Rossby numbers; we note that the outlier, which is saturated but has a small Rossby number, is a binary of Algol type.

Following Pizzolato et al. (2003) and Wright et al. (2011, 2018), we fit the relationship between the activity indicator $R_X = \log(F_X/F_{\text{bol}})$ and the Rossby number R_o by the ansatz

$$R_X = \begin{cases} R_{X,\text{sat}} & \log(R_o) < \log(R_{o,\text{sat}}) \\ \log(R_o)\cdot\beta + C & \log(R_o) \geq \log(R_{o,\text{sat}}), \end{cases} \quad (2)$$

where $R_{X,\text{sat}}$ is the saturation limit, $R_{o,\text{sat}}$ is the Rossby number at the saturation threshold to the saturation regime, β is the power-law index, and the constant is given by $C = R_{X,\text{sat}} - \log(R_{o,\text{sat}})\cdot\beta$. We applied the method of least squares to the detections in Fig. 9 excluding the outlier and found a saturation limit of $R_{X,\text{sat}} = -3.21 \pm 0.16$, a break value between saturated and unsaturated sources at $\log(R_{o,\text{sat}}) = -0.84 \pm 0.18$, and a power law index of $\beta = -2.12 \pm 0.60$. Unlike Wright et al. (2011), we did not explicitly exclude flares from our sample, nevertheless, the break value to the saturation regime and saturation limit of the X-ray activity from our Hyades sample is very similar to the values obtained by Wright et al. (2011). Therefore, we conclude again that the X-ray detections of the Hyades members are not strongly biased by flares. We find a power-law index that is slightly steeper than the canonical value of $\beta = -2$ but the canonical value and the value from Wright et al. (2011) are within the error margins.

4. Conclusion

In this paper we present an update on the X-ray and rotational properties of the Hyades. Using membership criteria based upon the recent *Gaia* DR2 release, we create a membership list containing 1066 main sequence Hyades stars both in the core and in the tidal tails of the Hyades cluster. Using all available X-ray data, we detect 281 Hyades members (~26%) as X-ray sources, which is a significant increase compared to 139 unique Hyades detections reported by Stelzer & Neuhäuser (2001) and 187 unique Hyades members reported by Stern et al. (1995); we note that due to differences in the Hyades membership list, the RASS data reduction, and the identification procedure, we only have 111 sources in common with Stern et al. (1995).

Confirming earlier results from Micela et al. (1988) and Stern et al. (1995), we find the highest detection fractions for F- and G-type stars, while the detection fraction decreases for K- and M-type members. We specifically detect Hyades members with X-ray luminosities in the range from $\sim 2 \times 10^{27}$ erg s^{-1} to $\sim 2 \times 10^{30}$ erg s^{-1} covering three orders of magnitude. The brightest sources are active binaries, while late M-type dwarfs reach the lowest X-ray luminosities. The X-ray luminosity distributions of F- and G-type members are much steeper than for K- and M-type dwarfs. Compared to Stern et al. (1995), we extend the distributions of the latest spectral types to lower luminosities. Since we do not find meaningful upper limits below 1×10^{27} erg s^{-1}, we expect that all Hyades members with a convection zone are actually detectable with sufficiently long *XMM-Newton* exposures. Furthermore, the upcoming eROSITA all-sky survey (Merloni et al. 2012, eRASS) is expected to provide nearly complete X-ray detections for the Hyades members, at least in the core. The observed X-ray activity ranges from $\log(F_X/F_{\rm bol}) \approx -5.7$ for early F-type sources and continuously increases for decreasing spectral type reaching the saturation limit at $\log(F_X/F_{\rm bol}) \approx -3$ for sources later than spectral type M3. Stars of types F and G show the smallest spread in their X-ray luminosities, however, the intrinsic spread increases for K- and M-type sources just like the spread in the observed rotation periods. For 104 Hyades members, we find multiple detections either by different instruments or in different pointings. Similarly to Stern et al. (1995) who compare X-ray luminosities from *Einstein* and RASS, we do not find strong variations attributable to solar-like activity cycles, although the observations have different time intervals with a maximum baseline of 25 years. Only a few Hyades members show variations larger than a factor of three and these sources are all M-type dwarfs, where flares are probably responsible for the variation.

For the first time, we analyze the X-ray properties of the Hyades tidal tails. The X-ray detection fraction is lower in the tails because of generally larger source distances. The X-ray luminosity distributions formally differ between the core and tails of the Hyades, however, this difference does not necessarily argue against their common origin but is likely caused by a higher contamination of field stars in the tail member sample. The now running eROSITA all-sky survey is expected to provide far more detections among the tidal tail members and will therefore allow a far more detailed comparison between core and tails.

We find rotation periods for 204 of our bona fide Hyades members, 103 of them (∼50%) are detected in X-rays, and Rossby numbers can be estimated. In an activity-rotation diagram, the linear increase of activity with decreasing Rossby number for slow rotators is visible as well as the saturation limit for the fast rotators. For the Hyades, sources with high and low Rossby numbers differ mainly by spectral type, however, the parameters of the activity-rotation relation are very similar to those obtained by Wright et al. (2011, 2018) who applied sources of different ages. The ongoing Transiting Exoplanet Survey Satellite (TESS) mission will significantly improve on the number of available rotation periods for Hyades members and thus again allow for a more detailed comparison between core and tails.

Acknowledgements. SF acknowledges supports through the Integrationsamt Hildesheim, the ZAV of Bundesagentur für Arbeit, and Hamburg University, JR acknowledges support by DLR under grant 50 QR 1605, and PCS through the SFB 676 founded by DFG and by DLR under grant 50 OR 1901. SF thanks Gabriele Uth and Maria Theresa Sangdaan-Lehmann for their support. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The paper is based on observations obtained with *XMM-Newton*, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA. This research has made use of data obtained from the *Chandra* Source Catalog, provided by the *Chandra* X-ray Center (CXC) as part of the *Chandra* Data Archive. This paper includes data collected by the K2 mission. Funding for the K2 mission was provided by the NASA Science Mission directorate. We used the VizieR catalog access tool and the SIMBAD database, operated at CDS, Strasbourg, France. The original description of the VizieR service was published in A&AS 143, 23.

References

Agrawal, P. C., Rao, A. R., & Sreekantan, B. V. 1986, MNRAS, 219, 225
Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103
Delorme, P., Collier Cameron, A., Hebb, L., et al. 2011, MNRAS, 413, 2218
Dempsey, R. C., Linsky, J. L., Fleming, T. A., & Schmitt, J. H. M. M. 1993, ApJS, 86, 599
Dempsey, R. C., Linsky, J. L., Fleming, T. A., & Schmitt, J. H. M. M. 1997, ApJ, 478, 358
Douglas, S. T., Agüeros, M. A., Covey, K. R., et al. 2016, ApJ, 822, 47
Douglas, S. T., Curtis, J. L., Agüeros, M. A., et al. 2019, ApJ, 879, 100
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37
Fleming, T. A., Liebert, J., Gioia, I. M., & Maccacaro, T. 1988, ApJ, 331, 958
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2018a, A&A, 616, A1
Gaia Collaboration (Babusiaux, C., et al.) 2018b, A&A, 616, A10
Hartman, J. D., Bakos, G. Á., Noyes, R. W., et al. 2011, AJ, 141, 166
Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Lanzafame, A. C., Distefano, E., Messina, S., et al. 2018, A&A, 616, A16
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
Lodieu, N., Smart, R. L., Pérez-Garrido, A., & Silvotti, R. 2019, A&A, 623, A35
Meingast, S., & Alves, J. 2019, A&A, 621, L3
Merloni, A., Predehl, P., Becker, W., et al. 2012, ArXiv e-prints [arXiv:`1209.3114`]
Micela, G., Sciortino, S., Vaiana, G. S., et al. 1988, ApJ, 325, 798
Micela, G., Sciortino, S., Kashyap, V., Harnden, F. R. J., & Rosner, R. 1996, ApJS, 102, 75
Pallavicini, R., Tagliaferri, G., & Stella, L. 1990, A&A, 228, 403
Pecaut, M. J., & Mamajek, E. E. 2013, ApJS, 208, 9
Peres, G., Orlando, S., Reale, F., Rosner, R., & Hudson, H. 2000, ApJ, 528, 537
Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., & Ventura, P. 2003, A&A, 397, 147
Pojmanski, G. 2002, Acta Astron., 52, 397
Prosser, C. F., Shetrone, M. D., Dasgupta, A., et al. 1995, PASP, 107, 211
Radick, R. R., Thompson, D. T., Lockwood, G. W., Duncan, D. K., & Baggett, W. E. 1987, ApJ, 321, 459
Radick, R. R., Lockwood, G. W., Skiff, B. A., & Thompson, D. T. 1995, ApJ, 452, 332
ROSAT, Consortium 2000, VizieR Online Data Catalog, IX/30
ROSAT Scientific Team 2000, VizieR Online Data Catalog, IX/28A
Röser, S., Schilbach, E., Piskunov, A. E., Kharchenko, N. V., & Scholz, R. D. 2011, A&A, 531, A92
Röser, S., Schilbach, E., & Goldman, B. 2019, A&A, 621, L2
Schmitt, J. H. M. M. 1997, A&A, 318, 215
Schröder, K.-P., Mittag, M., Jack, D., Rodriguez Jimenez, A., & Schmitt, J. H. M. M. 2020, MNRAS, 492, 1110
Stelzer, B., & Neuhäuser, R. 2001, A&A, 377, 538
Stelzer, B., Schmitt, J. H. M. M., Micela, G., & Liefke, C. 2006, A&A, 460, L35
Stern, R. A., Zolcinski, M. C., Antiochos, S. K., & Underwood, J. H. 1981, ApJ, 249, 647
Stern, R. A., Schmitt, J. H. M. M., & Kahabka, P. T. 1995, ApJ, 448, 683
Truemper, J. 1982, Adv. Space Res., 2, 241
Vilhu, O. 1984, A&A, 133, 117
Walter, F., Charles, P., & Bowyer, S. 1978, ApJ, 225, L119
Weisskopf, M. C., Brinkman, B., Canizares, C., et al. 2002, PASP, 114, 1
Worthey, G., & Lee, H.-C. 2011, ApJS, 193, 1
Wright, N. J., Drake, J. J., Mamajek, E. E., & Henry, G. W. 2011, ApJ, 743, 48
Wright, N. J., Newton, E. R., Williams, P. K. G., Drake, J. J., & Yadav, R. K. 2018, MNRAS, 479, 2351

Chapter 7

Conclusion and outlook

In this thesis, I presented methods to identify stellar coronal X-ray sources by a cross-correlation with optical counterparts mainly from *Gaia* and analyzed the obtained identifications. Due to the large sample size, this task cannot be carried out individually by hand, but rather an automatic algorithm was presented. The identifications allow to study complete, flux-limited samples of stellar X-ray sources. The coherent samples contain all types of stellar sources that were previously analyzed separately. To create sub-samples with a specific completeness and reliability for the desired scientific application, matching probabilities for every counterpart were derived. The described principles can also be applied to other identification tasks.

7.1 Summary

For the XMMSL catalog with a high positional accuracy and bright, less dense counterparts, identifications with a completeness and reliability of more than 96 % were obtained by only considering the angular separation between the X-ray source and selected optical counterparts. My collaborators and I found 5 920 bona fide stellar XMMSL2 sources, which is a significant improvement compared to previous identifications presented by Saxton et al. (2008). The exposure time of the XMMSL2 sources is quite short (typically 6 s), and therefore, many sources were detected in an active or flaring state. Hence, a substantial fraction of the stellar XMMSL2 sources were not detected in previous X-ray observations as the RASS.

For X-ray surveys with a large positional uncertainty and a high counterpart density, the probability of finding a possible association by chance in the searching radius of the X-ray source is high. To identify the correct association, further source properties in addition to the geometric match need to be considered. Since the X-ray luminosity of coronal sources saturates and stellar X-ray sources can only be detected in the solar vicinity, the X-ray to optical flux ratio, color, and counterpart distance are good properties to distinguish between true identifications and spurious associations.

This was applied by my coworkers and I to identify for the first time the whole stellar content of the RASS adopting a Bayesian framework. The identification procedure consists of a geometric crossmatch between RASS sources and eligible stellar counterparts. Furthermore, Bayes maps of the X-ray to optical flux ratio and the counterpart distances were applied to distinguish between likely true stellar and possibly random associations. I showed that the details of the estimation of the Bayes map have a minor influence on the results but the estimated probabilities strongly differ for Bayes maps derived for different X-ray surveys. Since the stellar counterparts substantially differ with Galactic latitude, varying counterpart densities, catalog fractions, and Bayes maps were adopted as a function of the Galactic position. Finally, 28 000 RASS sources are identified as stars with a completeness and reliability of about 93 %, which is the largest sample of stellar X-ray sources obtained so far. The sample contains stars of all spectral types. The onset of convection and the saturation limit are clearly visible. From the parallaxes of *Gaia* EDR3, the three dimensional distribution of the stellar RASS sources was obtained.

The identification procedure was also applied to the eFEDS field that provides the X-ray sensitivity expected after the completion of the *eROSITA* all-sky survey for about 140 deg^2. The results were compared to the identification obtained with a machine learning approach and another Bayesian algorithm that is not optimized for stellar sources. The identifications of all methods agree to about 90 %, which is the expected reliability of the stellar samples. The Bayesian procedure for stars was also applied to a test catalog. For this purpose, I simulated X-ray properties for sources of different type. The predicted completeness and reliability agrees well with the true values, again, confirming the reliability of the identification procedure.

The identification of the RASS sources, and further data from pointed observations of *ROSAT*, *Chandra*, and *XMM-Newton*, were applied to update the X-ray view of the Hyades cluster. My coworkers and I found X-ray detections for 281 of 1066 Hyades members adopting recent *Gaia* DR2 membership lists, which is a significant improvement compared to previous work. The X-ray to bolometric flux ratios of the 640 Myr old Hyades members show that dwarfs up to a spectral type of about M3 are still saturated, while the X-ray activity constantly decreases for earlier type sources. The X-ray luminosity of most multiply detected Hyades members differ by less than a factor of three, and these variations are likely caused by flares and not by solar-like activity cycles. The activity-rotation relation described by Wright et al. (2011, 2018) is confirmed by the Hyades members although the Rossby numbers for these source are a function of spectral type rather than age.

7.2 Future applications and improvements

The currently ongoing *eROSITA* all-sky survey is expected to be completed at the end of the year 2023 and about 800 000 stars are expected to be detected. The stellar identifications of soft X-ray surveys presented in this thesis provide valuable experiences for different aspects of the identification method needed for eRASS. The XMMSL2 sources cover a similar spectral range and also the positional accuracies are similar. However, the XMMSL2 catalog is much shallower and the scanning law is different. On the other hand, the eFEDS catalog provides a similar sensitivity as expected at the end of eRASS:8 but it covers a rather small part of the sky. Therefore, the eFEDS identifications are appropriate to test the properties that are best suited to identify the stellar counterpart. However, the change of these parameters with sky position can be analyzed only with the RASS, which has a very similar scanning law as eRASS. On the other hand, the RASS is less sensitive and the positional uncertainties are much larger. The data from the first four eRASS surveys are already available. With a similar procedure as presented in this thesis, I already obtained preliminary stellar identifications of the first *eROSITA* all-sky survey and found more than 100 000 stars only in the half of the sky with Galactic latitudes $l > 180°$.

However, due to the deep X-ray exposure of eRASS, optically faint counterparts with a high source density, especially in the Galactic plane, are expected to be detected. Therefore, the identification of the correct association will be more difficult compared to RASS, although the positional accuracy of the eRASS sources is much higher. Hence, additional information need to be adopted. For the identifications described in this thesis, only two properties in addition to the geometric match are applied but there is theoretically no limit of parameters in the Bayesian framework. In practice, the training set has to be large enough to cover the n-dimensional parameter space and to derive meaningful Bayes maps. Due to the large size of the eRASS catalog and the high positional accuracy, much more training set sources with a good positional match are available, and therefore, more properties can be considered.

Most coronal sources are known to be quite soft X-ray emitters. Therefore, the hardness ratio provides additional information on the reliability of the counterparts. Associations detected at a high X-ray to bolometric flux ratio are unlikely coronal emitters if the X-ray flux represents the quiescent value, but the identifications might be reasonable if the source is detected during a flare. During the eight all-sky surveys with *eROSITA*, every source will be observed multiple times, which allows an estimation of the variability and the quiescent flux of the eRASS sources. With this information at hand, the identification of flaring sources will be improved. For future *Gaia* releases, chromospheric activity indicators estimated from spectra in the range of the Ca II triplet will improve the reliability at least for some of the potential associations. First spectra for a few sources will be already published in the second quarter of the year 2022 as part of *Gaia* DR3. Furthermore, the Transiting Exoplanet Survey Satellite (TESS) will provide rotation period for some counterparts. The agreement between the rotation period and X-ray activity will help to further constrain reasonable associations.

Further improvements of the identification algorithm are planed. A limiting factor of the current procedure is the size of the training set. In future applications, the properties of the true identifications can also be obtained by the difference between the property distribution of counterparts to true and randomized X-ray sources. However, this approach needs to be tested properly. The Bayes map strongly depends on the adopted X-ray sample because the automatic procedure is optimized for an average source. Therefore, the brightest sources are not weighted properly, a problem which is especially sensitive in the case of the eRASS catalog that covers a large range of X-ray fluxes. This problem can be reduced by applying an invariant Bayes map that is estimated, for example, from X-ray to optical flux ratios of the true identifications and the optical fluxes of the random associations.

With the outlined improvements, I am very confident that the stellar content of eRASS can be identified with a high reliability. With the unprecedented number of coronal X-ray emitters, we will be able to investigate well characterized large subsamples of specific stellar sources. The identified stellar X-ray sources can be used to investigate the X-ray properties of open clusters and star forming regions. For example, Schmitt et al. (2021) adopted preliminary identifications of eRASS1 sources to reveal and analyze the low-mass population of the Sco-Cen OB association. Furthermore, virtually all Hyades members are expected to be detected in eRASS. Therefore, X-ray emission will be an important criterion for the membership classification. By comparing the X-ray activity of members from stellar clusters with different ages, the activity-age relation will be examined in an unprecedented way with the stellar eRASS identifications. In conjunction with rotation periods from TESS, also the activity-rotation relation will be investigated in unprecedented detail. The identification of the stellar eRASS sources will therefore enable us to utilize the groundbreaking potential of *eROSITA* to stellar X-ray science.

Bibliography

Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8

Audard, M., Güdel, M., Drake, J. J., & Kashyap, V. L. 2000, ApJ, 541, 396

Baliunas, S. L., Donahue, R. A., Soon, W. H., et al. 1995, ApJ, 438, 269

Barnes, S. A. 2003, ApJ, 586, 464

Bayes, T. & Price. 1763, Philosophical Transactions of the Royal Society of London, 53, 370

Berdyugina, S. V., Berdyugin, A. V., Ilyin, I., & Tuominen, I. 1998, A&A, 340, 437

Berghoefer, T. W., Schmitt, J. H. M. M., & Cassinelli, J. P. 1996, A&AS, 118, 481

Berghoefer, T. W., Schmitt, J. H. M. M., Danner, R., & Cassinelli, J. P. 1997, A&A, 322, 167

Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103

Briel, U. G., Aschenbach, B., Hasinger, G., et al. 1996, The ROSAT Users' Handbook

Brunner, H., Liu, T., Lamer, G., et al. 2021, arXiv e-prints, arXiv:2106.14517

Budavári, T. & Szalay, A. S. 2008, ApJ, 679, 301

Burnight, T. R. 1949, Phys. Rev., 76, 165

Catura, R. C., Acton, L. W., & Johnson, H. M. 1975, ApJL, 196, L47

Dempsey, R. C., Linsky, J. L., Fleming, T. A., & Schmitt, J. H. M. M. 1993, ApJS, 86, 599

den Herder, J. W., Brinkman, A. C., Kahn, S. M., et al. 2001, A&A, 365, L7

Edlén, B. 1943, ZA, 22, 30

Evans, D. W., Riello, M., De Angeli, F., et al. 2017, A&A, 600, A51

Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4

Fabricius, C., Luri, X., Arenou, F., et al. 2021, A&A, 649, A5

Fleming, T. A., Giampapa, M. S., Schmitt, J. H. M. M., & Bookbinder, J. A. 1993, ApJ, 410, 387

Fleming, T. A., Schmitt, J. H. M. M., & Giampapa, M. S. 1995, ApJ, 450, 401

Freund, S., Robrade, J., Schneider, P. C., & Schmitt, J. H. M. M. 2018, A&A, 614, A125

Freund, S., Robrade, J., Schneider, P. C., & Schmitt, J. H. M. M. 2020, A&A, 640, A66

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, A&A, 616, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016a, A&A, 595, A2

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016b, A&A, 595, A1

Gaia Collaboration, Spoto, F., Tanga, P., et al. 2018b, A&A, 616, A13

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2013, Bayesian Data Analysis, 3rd edn. (Chapman and Hall/CRC)

Giacconi, R., Branduardi, G., Briel, U., et al. 1979, ApJ, 230, 540

Giacconi, R., Gursky, H., Paolini, F. R., & Rossi, B. B. 1962, PhRvL, 9, 439

Giacconi, R., Kellogg, E., Gorenstein, P., Gursky, H., & Tananbaum, H. 1971, ApJL, 165, L27

Gilman, P. A. 1983, ApJS, 53, 243

Gioia, I. M., Maccacaro, T., Schild, R. E., et al. 1990, ApJS, 72, 567

Gondoin, P. 1999, A&A, 352, 217

Grotrian, W. 1939, Naturwissenschaften, 27, 214

Güdel, M. 2004, A&A Rv, 12, 71

Güdel, M., Audard, M., Reale, F., Skinner, S. L., & Linsky, J. L. 2004, A&A, 416, 713

Guedel, M., Schmitt, J. H. M. M., Benz, A. O., & Elias, N. M., I. 1995, A&A, 301, 201

Guillout, P., Schmitt, J. H. M. M., Egret, D., et al. 1999, A&A, 351, 1003

Haakonsen, C. B. & Rutledge, R. E. 2009, ApJS, 184, 138

Habbal, S. R., Druckmüller, M., Alzate, N., et al. 2021, ApJL, 911, L4

Hall, J. C. 2008, Living Reviews in Solar Physics, 5, 2

Hempelmann, A., Schmitt, J. H. M. M., Schultz, M., Ruediger, G., & Stepien, K. 1995, A&A, 294, 515

Holl, B., Audard, M., Nienartowicz, K., et al. 2018, A&A, 618, A30

Huensch, M., Schmitt, J. H. M. M., Schroeder, K.-P., & Reimers, D. 1996, A&A, 310, 801

Huensch, M., Schmitt, J. H. M. M., & Voges, W. 1998a, A&AS, 127, 251

Huensch, M., Schmitt, J. H. M. M., & Voges, W. 1998b, A&AS, 132, 155

Hussain, G. A. J., Jardine, M., Donati, J. F., et al. 2007, MNRAS, 377, 1488

Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1

Jardine, M. 2004, A&A, 414, L5

Kashyap, V. L., Drake, J. J., Güdel, M., & Audard, M. 2002, ApJ, 580, 1118

Katsova, M. M., Drake, J. J., & Livshits, M. A. 1999, ApJ, 510, 986

Katz, D., Sartoretti, P., Cropper, M., et al. 2019, A&A, 622, A205

LaPlace, P. S. 1814, Essai philosophique sur les probabilités (Courcier)

Lin, R. P., Schwartz, R. A., Kane, S. R., Pelling, R. M., & Hurley, K. C. 1984, ApJ, 283, 421

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2

Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, 595, A4

Linsky, J. L. & Haisch, B. M. 1979, ApJL, 229, L27

Loredo, T. J. 1992, in Statistical Challenges in Modern Astronomy, 275–297

Maggio, A., Vaiana, G. S., Haisch, B. M., et al. 1990, ApJ, 348, 253

McGale, P. A., Pye, J. P., & Hodgkin, S. T. 1996, MNRAS, 280, 627

Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv e-prints, arXiv:1209.3114

Micela, G., Sciortino, S., Vaiana, G. S., et al. 1988, ApJ, 325, 798

Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. 1984a, ApJ, 279, 763

Noyes, R. W., Weiss, N. O., & Vaughan, A. H. 1984b, ApJ, 287, 769

Pallavicini, R., Golub, L., Rosner, R., et al. 1981, ApJ, 248, 279

Pallavicini, R., Serio, S., & Vaiana, G. S. 1977, ApJ, 216, 108

Parker, E. N. 1955, ApJ, 122, 293

Parker, E. N. 1979, Cosmical magnetic fields. Their origin and their activity

Parker, E. N. 1988, ApJ, 330, 474

Parzen, E. 1962, The Annals of Mathematical Statistics, 33, 1065

Patten, B. M. & Simon, T. 1996, ApJS, 106, 489

Peres, G., Orlando, S., Reale, F., Rosner, R., & Hudson, H. 2000, ApJ, 528, 537

Pevtsov, A. A., Fisher, G. H., Acton, L. W., et al. 2003, ApJ, 598, 1387

Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., & Ventura, P. 2003, A&A, 397, 147

Pizzolato, N., Maggio, A., & Sciortino, S. 2000, A&A, 361, 614

Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1

Randich, S., Schmitt, J. H. M. M., Prosser, C. F., & Stauffer, J. R. 1996, A&A, 305, 785

Reimers, D., Huensch, M., Schmitt, J. H. M. M., & Toussaint, F. 1996, A&A, 310, 813

Riello, M., De Angeli, F., Evans, D. W., et al. 2018, A&A, 616, A3

Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3

Robrade, J., Schmitt, J. H. M. M., & Favata, F. 2012, A&A, 543, A84

Rosenblatt, M. 1956, The Annals of Mathematical Statistics, 27, 832

Rutledge, R. E., Brunner, R. J., Prince, T. A., & Lonsdale, C. 2000, ApJS, 131, 335

Salvato, M., Buchner, J., Budavári, T., et al. 2018, MNRAS, 473, 4937

Salvato, M., Wolf, J., Dwelly, T., et al. 2021, arXiv e-prints, arXiv:2106.14520

Saxton, R. D., Read, A. M., Esquej, P., et al. 2008, A&A, 480, 611

Schmitt, J. H. M. M. 1997, A&A, 318, 215

Schmitt, J. H. M. M., Czesla, S., Freund, S., Robrade, J., & Schneider, P. C. 2021, arXiv e-prints, arXiv:2106.14549

Schmitt, J. H. M. M., Fleming, T. A., & Giampapa, M. S. 1995, ApJ, 450, 392

Schmitt, J. H. M. M., Golub, L., Harnden, F. R., J., et al. 1985, ApJ, 290, 307

Schmitt, J. H. M. M. & Liefke, C. 2004, A&A, 417, 651

Schneider, P. C., Freund, S., Czesla, S., et al. 2021, arXiv e-prints, arXiv:2106.14521

Schrijver, C. J. & Zwaan, C. 2008, Solar and Stellar Magnetic Activity

Schwarzschild, M. 1948, ApJ, 107, 1

Scott, D. 1992, Multivariate Density Estimation: Theory, Practice, and Visualization, A Wiley-interscience publication (Wiley)

Secrest, N. J., Dudik, R. P., Dorland, B. N., et al. 2015, ApJS, 221, 12

Skumanich, A. 1972, ApJ, 171, 565

Stępień, K., Schmitt, J. H. M. M., & Voges, W. 2001, A&A, 370, 157

Stelzer, B. & Neuhäuser, R. 2001, A&A, 377, 538

Stelzer, B., Schmitt, J. H. M. M., Micela, G., & Liefke, C. 2006, A&A, 460, L35

Stern, R. A., Schmitt, J. H. M. M., & Kahabka, P. T. 1995, ApJ, 448, 683

Stocke, J. T., Morris, S. L., Gioia, I. M., et al. 1991, ApJS, 76, 813

Strüder, L., Briel, U., Dennerl, K., et al. 2001, A&A, 365, L18

Testa, P., Saar, S. H., & Drake, J. J. 2015, Philosophical Transactions of the Royal Society of London Series A, 373, 20140259

Traulsen, I., Schwope, A. D., Lamer, G., et al. 2020, A&A, 641, A137

Trümper, J. 1984, Physica Scripta Volume T, 7, 209

Turner, M. J. L., Abbey, A., Arnaud, M., et al. 2001, A&A, 365, L27

Vaiana, G. S., Cassinelli, J. P., Fabbiano, G., et al. 1981, ApJ, 245, 163

Vaiana, G. S., Krieger, A. S., & Timothy, A. F. 1973, SoPh, 32, 81

van Ballegooijen, A. A., Asgari-Targhi, M., Cranmer, S. R., & DeLuca, E. E. 2011, ApJ, 736, 3

Vilhu, O. 1984, A&A, 133, 117

Vilhu, O. & Rucinski, S. M. 1983, A&A, 127, 5

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

Walter, F. M. & Bowyer, S. 1981, ApJ, 245, 671

Weisskopf, M. C., Brinkman, B., Canizares, C., et al. 2002, PASP, 114, 1

Wolter, H. 1952, Annalen der Physik, 445, 94

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

Wright, N. J., Drake, J. J., Mamajek, E. E., & Henry, G. W. 2011, ApJ, 743, 48

Wright, N. J., Newton, E. R., Williams, P. K. G., Drake, J. J., & Yadav, R. K. 2018, MNRAS, 479, 2351

Acknowledgments

My PhD would not have been possible without the support of many people, whom I would like to thank.

First of all, I thank Jürgen Schmitt for his great supervision, providing many helpful ideas, and shearing his great knowledge. I also thank him for his support in making my PhD possible.

Furthermore, I thank my colleagues, especially Stefan Czesla, Jan Robrade, and Christian Schneider, for their advice and guiding me through the different challenges of a PhD.

I thank Gabriele Uth and Maria-Theresa Lehmann for their support and the good company during the time of my PhD and hopefully many more years to come.

I thank Manuel Güdel that he agreed to compose the second assessment of this doctoral dissertation, and also Jochen Liske, Marcus Brüggen, and Peter Hauschildt for serving in my examination commission.

I acknowledge financial support from the Hamburg University, the Bundesagentur für Arbeit, the Integrationsamt of Hamburg and Hildesheim, and the Landkreis Uelzen.

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Ort, Datum

Unterschrift des Doktoranden