

Cross-correlation of the 2XMMi catalogue with Data Release 7 of the Sloan Digital Sky Survey[★]

F.-X. Pineau¹, C. Motch¹, F. Carrera², R. Della Ceca³, S. Derrière¹, L. Michel¹, A. Schwoppe⁴, and M. G. Watson⁵

¹ CNRS, Université de Strasbourg, Observatoire Astronomique, 11 rue de l'Université, 67000 Strasbourg, France
e-mail: francois-xavier.pineau@astro.unistra.fr

² Instituto de Física de Cantabria (CSIC-UC), 39005 Santander, Spain

³ INAF - Osservatorio Astronomico di Brera, via Brera 28, 20121 Milan, Italy

⁴ Astrophysikalisches Institut Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany

⁵ Department of Physics and Astronomy, University of Leicester, LE1 7RH, UK

Received 2 Juin 2010 / Accepted 22 November 2010

ABSTRACT

The Survey Science Centre of the XMM-Newton satellite released the first incremental version of the 2XMM catalogue in August 2008. Containing more than 220 000 X-ray sources, the 2XMMi was at that time the largest catalogue of X-ray sources ever published and thus constitutes an unprecedented resource for studying the high-energy properties of various classes of X-ray emitters such as AGN and stars. Thanks to the high throughput of the EPIC cameras on board XMM-Newton accurate positions, fluxes, and hardness ratios are available for a substantial fraction of the X-ray detections. The advent of the 7th release of the Sloan Digital Sky Survey offers the opportunity to cross-match two major surveys and extend the spectral energy distribution of many 2XMMi sources towards the optical bands. This implies building extensive homogeneous samples with a statistically controlled rate of spurious matches and completeness. We here present a cross-matching algorithm based on the classical likelihood ratio estimator. The method developed has the advantage of providing true probabilities of identifications without resorting to heavy Monte-Carlo simulations. Over 30,000 2XMMi sources have SDSS counterparts with individual probabilities of identification higher than 90%. At this threshold, the sample has only 2% spurious matches and contains 77% of all expected SDSS identifications. Using spectroscopic identifications from the SDSS DR7 catalogue supplemented by extraction from other catalogues, we build an identified sample from which the way the various classes of X-ray emitters gather in the multi dimensional parameter space can be analysed and later used to design a source classification scheme. We illustrate the interest of this clean source sample by investigating two scientific use cases. In the first example we show how these multi-wavelength data can be used to search for new QSO2s. Although no specific range of observed properties allows us to efficiently identify Compton Thick QSO2s, we show that the prospects are much better for Compton Thin AGN2 and discuss several possible multi-parameter selection strategies. In a second example, we confirm the hardening of the mean X-ray spectrum with increasing X-ray luminosity on a sample of over 500 X-ray active stars and reveal that on average X-ray active M stars display bluer $g - r$ colour indexes than less active ones. Although this catalogue of 2XMM-SDSS sources cannot be used directly for statistical studies, it nevertheless represents an excellent starting point to select well defined samples of X-ray-emitting objects.

Key words. methods: statistical – catalogs – X-rays: general – stars: activity – quasars: general

1. Introduction

The growing collecting area and sensitivity of modern astronomical detectors combined with the increasing storage and processing capabilities offered by current computer facilities has made possible the gathering on comparatively short time scales of very large sky surveys that were beyond reach only a few years ago. Most parts of the electromagnetic spectrum benefit from this evolution. Among recently completed or ongoing projects are the Two Micron All Sky Survey (2MASS) (Cutri et al. 2003) and the Sloan Digital Sky Survey (Adelman-McCarthy et al. 2008) for instance. Space-borne missions currently in operation

such as the *Spitzer* Space Telescope (Werner et al. 2004) observing in the infra-red or the Chandra (Weisskopf et al. 2000) and XMM-Newton (Jansen et al. 2001) X-ray observatories are collecting at a high rate a wealth of measurements on an unprecedented number of objects in their energy range. In the relatively near future, ground-based automated very large telescopes such as pan-STARRS (Wang et al. 2010) or such as the Large Synoptic Survey Telescope (Tyson 2002) will collect detailed photometric information on a breathtaking number of faint galaxies.

Merging measurements arising from several instruments allows us to build spectral energy distributions in a range of wavelengths extending over a large part of the electromagnetic spectrum. The recent availability of wide angle surveys with high detection sensitivities allows us to measure with comparable accuracies and in several scientifically important wavelength ranges the spectral energy density of the main classes of X-ray emitting astrophysical sources. Building large homogeneous samples provides valuable insight on the emission mechanisms

[★] The corresponding fits file can be downloaded from the XCat-DB home page (<http://xcatdb.u-strasbg.fr/>). The file also contains line information for all SDSS spectroscopic entries matching a 2XMM source. Results from the cross-correlation with the 2XMM DR3 are also available at the same location. The 2XMMi/SDSS DR7 cross-correlation data file is also available at the CDS via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/527/A126>

and evolutionary processes and may allow the detection of rare objects or outliers, which would be otherwise hard to unveil in smaller samples. In this respect, a good estimate of the true rate of false cross-identification is important to assess the relevance of any group of outliers.

However, the gathering of large groups of sources with well characterised multi-wavelength properties first requires a proper handling of the cross-matching process between two or more catalogues. Although spatial resolution at high-energy steadily increased during the last years and may go on improving in the future, source density also grows as a result of the improved sensitivity, and the risk of confusion between unrelated objects detected at different wavelengths does not necessarily vanish. The confusion problem can be particularly arduous when comparing catalogues with very different spatial resolutions and densities, a problem often encountered in the identification process of high-energy sources which in several cases lack the superb spatial resolution affordable for instance in the optical domain, see e.g. Rutledge et al. (2000) for the identification of ROSAT sources and Luo et al. (2010) for a recent example involving multi-wavelength catalogues with different depths and angular resolutions.

The XMM-Newton satellite (Jansen et al. 2001) was launched by the European Space Agency late in 1999. XMM-Newton is currently the X-ray (0.2–12 keV) telescope in operation with the largest effective area. Three co-aligned telescopes feed two EPIC MOS (Turner et al. 2001) and one EPIC pn (Strüder et al. 2001) cameras. Two reflection grating arrays deviate about half of the X-ray photons from the EPIC MOS camera towards two Reflection Grating Spectrometers (RGS; den Herder et al. 2001). An optical monitor (OM; Mason et al. 2001), providing UV and optical images of a fraction of the field of view covered by the EPIC cameras down to the 21st mag, complements the X-ray instrumentation. One of the remarkable properties offered by the X-ray telescopes on-board XMM-Newton is to provide a large field of view of 30' diameter with a weakly degraded image point-spread function and low vignetting even at large off-axis angles. Accordingly, a large number of sources may be serendipitously discovered around the main target of the observation, which builds up to make an X-ray survey with an unprecedented combination of sensitivity and area covered. Starting from the beginning of the project, ESA recognised the high scientific interest of exploiting the XMM-Newton survey and appointed the present Survey Science Centre (SSC) on a competitive basis. Lead by the University of Leicester, the SSC is a consortium of ten European institutes conducting its activity on behalf of ESA. The SSC responsibilities have been presented in Watson et al. (2001). One of the most demanding tasks given to the consortium is the compilation of a catalogue of all sources serendipitously discovered in the field of view of the X-ray instruments and of their characterisation and identification at least in a statistical way.

Several spectroscopic identification campaigns and multi-wavelength studies have been recently performed by the SSC on samples of thousands of EPIC sources using follow-up observations at 4-m and 8-m class telescopes. The availabilities of the recently published SDSS Data Release 7 (DR7) and of the incremental version of the 2XMM catalogue (2XMMi) offer a unique opportunity to extend the identification work to a much more extended sky area. With its spectroscopic and photometric limiting magnitude about 2 mag brighter than that typically reached for the SSC source samples, SDSS identifications of XMM-Newton sources conveniently expand the identified sample towards brighter magnitudes and at the same time provide

access to a rich group of accurately quantified photometric and spectroscopic data.

As part of its scientific activities, the Survey Science Centre of the XMM-Newton satellite has developed a specific cross-correlation algorithm yielding actual probabilities of identification based on positional coincidence and applied this algorithm to the cross-identification of the 2XMMi and SDSS DR7 catalogues, thus creating one of the largest set of optically identified X-ray sources available so far. The result of the cross-correlation is made available as a separate fits file and is also available through the XCat-DB¹ (Motch et al. 2007; Michel et al. 2009).

The first sections of this paper present the details of the algorithm used to identify 2XMMi X-ray sources with SDSS DR7 optical objects. We apply the commonly used likelihood ratio to quantify the chance that a SDSS object is the counterpart of the X-ray source. Identification probabilities are computed with an original method that does not rely on Monte Carlo simulations and thus offers a better efficiency when cross-correlating large sets of data. We then describe the range of optical and X-ray parameters occupied by the main astrophysical classes of X-ray emitters and show how source classification could be achieved on this basis. In the last part of this paper, we investigate two example science cases, the search for new QSO2s, and the study of the properties of the X-ray active late-type star population.

2. Description of the cross-correlated catalogues

2.1. 2XMMi catalogue

The incremental Second XMM-Newton Serendipitous Source Catalogue (2XMMi) is an extended version of the 2XMM Catalogue (Watson et al. 2009). It has been built from 4117 individual pointed observations performed by the XMM-Newton Observatory and contains 289 083 heterogeneous detections for a total of 221 012 unique X-ray sources. The catalogue covers ~1% of the sky over a large range of Galactic latitudes and longitudes. Owing to the wide range of exposure times, the area covered sensitively depends on limiting flux and energy range (see Fig. 8 in Watson et al. 2009). A 90% complete relative sky coverage is reached at $F_X = 1$ and 9×10^{-14} erg cm⁻² s⁻¹ in the 0.5–2.0 keV and 2.0–12.0 keV bands respectively. The EPIC cameras encompass a field (FOV) of ~30' diameter and are sensitive in the energy range of ~0.2–12 keV. Source positions have a typical accuracy of ~2". In this paper, we limit our analysis to point-like sources with a positional error smaller or equal to 5". A source is defined as point-like if its extent maximum likelihood parameter (ep_ext_ml) is ≤ 4 . The resulting 2XMMi source sample consists of 264,361 detections and 200,067 unique 2XMMi sources.

2.2. SDSS Data Release 7

The Seventh Data Release of the Sloan Digital Sky Survey (Abazajian et al. 2009), covers 11663 deg², mostly in the northern Galactic cap. A total of 357 million objects have 5 band photometry, among which 1.6 million galaxies, quasars, and stars were spectroscopically observed. Most of the ~2000 deg² increment over data release 6 are located at low galactic latitude. Astrometric errors are <0.1" rms. At the 3% error level, the catalogue reaches magnitude limits in the range of 20.5 to 22.2 in the five photometric bands – *u*, *g*, *r*, *i* and *z* –. In this paper we only consider the so-called primary sources of the SDSS

¹ <http://xcatdb.u-strasbg.fr/>

DR7 Photometric Catalogue as available from the VizieR data server. Primary sources are the “main” detection of an object and have the best defined set of parameters. For most scientific applications, the primary detections are the only ones needed. Source lists have been extracted using the VO ConeSearch protocol. The central point of each query is the centre of the FOV of the XMM-Newton observation considered and the search radius is the distance from the centre to the farthest X-ray source, to which we add 3' for completeness.

3. Counterpart identification procedure

We discuss in (3.1) how we select optical candidates, taking into account arbitrary error ellipses on the source’s spherical coordinates. We compute a likelihood ratio (LR) for each target-candidate pair (3.2). This LR involves a measure of the local density using a kernel smoothing method (Appendix B). Estimating the true LR distribution for spurious associations (3.3) then allows us to compute for each target-candidate pair the probability of association only based on positional coincidence (3.4).

3.1. Selection of optical candidates

3.1.1. Selection criterion

We consider a target X-ray source and a candidate optical source with α_X, δ_X the equatorial coordinates of the X-ray source; $\sigma_{\alpha_X}, \sigma_{\delta_X}$ and $\rho_{\alpha,\delta}$ the error on $\alpha_X \cos \delta_X$ and on δ_X and the correlation between σ_{α_X} and σ_{δ_X} respectively; α_o, δ_o the equatorial coordinates of an optical source $\sigma_{\alpha_o}, \sigma_{\delta_o}$ and $\rho_{\alpha,\delta}$ the error on $\alpha_o \cos \delta_o$ and on δ_o and the correlation between σ_{α_o} and σ_{δ_o} , respectively.

As everybody implicitly does – except Budavári & Szalay (2008) –, we convert the spherical problem into a plane one and positional errors are interpreted as usual 2D Gaussians. We have chosen a projection on a 2D plane with a frame centred on the position of the X-ray source and having for x -axis the direction of the optical candidate (Fig. 1). Errors on positions become Gaussians: $\mathcal{N}_X(x, y; \sigma_{x_X}^2, \sigma_{y_X}^2, \rho_X \sigma_{x_X} \sigma_{y_X})$ and $\mathcal{N}_o(x - d, y; \sigma_{x_o}^2, \sigma_{y_o}^2, \rho_o \sigma_{x_o} \sigma_{y_o})$ with d the angular distance between the X-ray and the optical source. As suggested by Sinnott (1984), d is computed using the Haversine function. The transformation of the ellipses in the new reference frame is described in Appendix A.

The density of probability that the two sources are at the same location, and thus are the same object, is given by the convolution product of these two distributions. It leads to a new Gaussian:

$$P(x, y) = \mathcal{N}_c(x, y; \sigma_{x_c}, \sigma_{y_c}, \rho_c \sigma_{x_c} \sigma_{y_c}), \quad (1)$$

with $\sigma_{x_c}^2 = \sigma_{x_X}^2 + \sigma_{x_o}^2$, $\sigma_{y_c}^2 = \sigma_{y_X}^2 + \sigma_{y_o}^2$ and $\rho_c \sigma_{x_c} \sigma_{y_c} = \rho_X \sigma_{x_X} \sigma_{y_X} + \rho_o \sigma_{x_o} \sigma_{y_o}$.

If the optical source is the counterpart of the X-ray source, it falls with a probability γ inside the ellipse defined by the equation

$$\begin{pmatrix} x \\ y \end{pmatrix}^t \begin{pmatrix} \sigma_{x_c}^2 & \rho_c \sigma_{x_c} \sigma_{y_c} \\ \rho_c \sigma_{x_c} \sigma_{y_c} & \sigma_{y_c}^2 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = k_\gamma^2. \quad (2)$$

The completeness we have chosen is a 3σ criterion, often used as a compromise between the total number of associations and the number of counterparts missed (0.3%). This completeness, $\gamma = 99.7\%$, leads in 2D to $k_\gamma = 3.43935$. In the frame we have chosen, the coordinates of the optical source are $x = d$ and

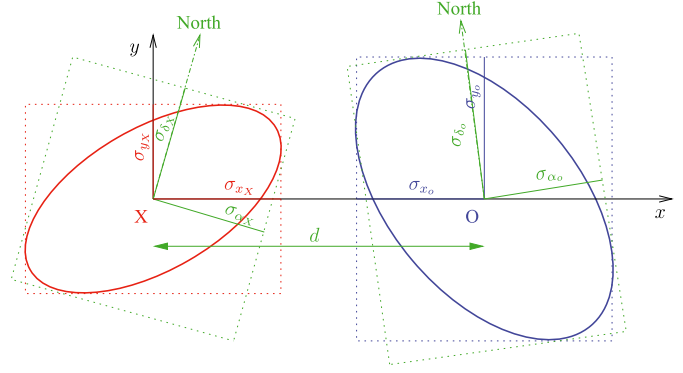


Fig. 1. Chosen projection plane: the xy frame is centred on the X-ray source position X; the x -axis is the direction towards the optical candidate, located at point O. d is the angular distance between the two sources. This frame is useful at high declinations when we cannot consider the meridians any longer – the directions of the north pole in X and in O – to be parallel. It allows us to deal naturally with the poles.

$y = 0$. The selection criterion we adopt will retain all candidates satisfying

$$\frac{d}{\sigma_{x_c} \sqrt{1 - (\rho_c \sigma_{x_c} \sigma_{y_c})^2}} \leq k_\gamma. \quad (3)$$

We make the additional following hypotheses. First, we neglect any systematic offset between the positions of the two catalogues. The 2XMMi catalogue as a whole is free of any systematic positional offset in a direction of the sky. This has been checked by cross-correlating the 2XMM catalogue with the SDSS DR5 Quasar catalogue (Watson et al. 2009). For a large number of cases (74% at $|b| > 20^\circ$), it was possible to correct the astrometry by cross-correlating field X-ray sources with USNO B1.0 entries. When no reliable astrometric correction could be found, increasing the applied systematic error from $0.35''$ to $1.0''$ accounts for the possible remaining coordinate offset and rotation affecting all the sources detected in a given observation (Watson et al. 2009). Second, we assume that all positions and associated errors have been computed at the same epoch and therefore corrected for proper motions.

3.1.2. Application to XMM-SDSS DR7 data

The 2XMMi catalogue provides a circular error on position ($radec_err$) and a systematic error ($syserr$) for each source. The error on positions of each X-ray source is the quadratic sum of these two values:

$$\sigma_{\alpha_X} = \sigma_{\delta_X} = \sqrt{radec_err^2 + syserr^2}. \quad (4)$$

Because it is symmetric, we have $\rho_{\alpha,\delta} = 0$, $\rho_X = 0$ and thus $\rho_c \sigma_{x_c} \sigma_{y_c}$ is directly equal to $\rho_o \sigma_{x_o} \sigma_{y_o}$.

Positional errors are elliptical in the SDSS DR7 catalogue: $\sigma_{\alpha_o} = raErr$, $\sigma_{\delta_o} = decErr$ and $\rho_{\alpha,\delta} = raDecCorr$. The definitions of the different parameters are summarized in Table 1.

3.2. Likelihood ratio

We compute a likelihood ratio (LR) for each target-candidate pair meeting the criterion of Eq. (3): the probability of finding the optical counterpart at a normalised distance r (see below) divided by the probability of having a spurious object at that distance.

Table 1. Summary of the astrometric parameters for the 2XMM catalogue and for the SDSS DR7.

	2XMMi	SDSS DR7
σ_α	$\sqrt{\text{radec_err}^2 + \text{syserr}^2}$	<i>raErr</i>
σ_δ	$\sqrt{\text{radec_err}^2 + \text{syserr}^2}$	<i>decErr</i>
$\rho_{\alpha,\delta}$	0	<i>raDecCorr</i>

The density of probability that the two sources are at the same location knowing x and y corresponds to the density of probability of having the counterpart in x and y , assuming that it is the same astrophysical object as the X-ray emitting one. The Gaussian \mathcal{N}_c (Eq. (1)) can be written in its canonical form $\frac{1}{2\pi\sigma_M\sigma_m} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_M^2} + \frac{y^2}{\sigma_m^2}\right)\right]$. Where σ_M and σ_m are the semi-major and semi-minor axis, in the eigenvector frame (x_1, y_1) , given by the eigendecomposition of the variance-covariance matrix of \mathcal{N}_c ,

$$\sigma_{M,m} = \frac{1}{2}(\sigma_{x_c}^2 + \sigma_{y_c}^2 \pm \sqrt{(\sigma_{x_c}^2 - \sigma_{y_c}^2)^2 + 4(\rho\sigma_{x_c}\sigma_{y_c})^2}). \quad (5)$$

We change the scale and switch to polar coordinates, which leads to the dimensionless Rayleigh distribution:

$$r = \sqrt{\frac{x_1^2}{\sigma_M^2} + \frac{y_1^2}{\sigma_m^2}}. \quad (6)$$

Therefore, the new elementary surface becomes $\pi\sigma_M\sigma_m$, the surface of the 1σ (or $r = 1$) ellipse.

The LR we use is inspired by the one described in [De Ruiter et al. \(1977\)](#). As [Wolstencroft et al. \(1986\)](#), we do not only consider the first candidate, but all sources satisfying Eq. (3). We thus replace the probability “of finding the first confusing object at a distance lying between r and $r + dr$ ” by the one of finding a confusing object between r and $r + dr$.

The probability of finding the optical counterpart (cp) at a distance lying between r and $r + dr$ is

$$dp(r|cp) = re^{-\frac{1}{2}r^2} dr. \quad (7)$$

And the probability of finding a spurious object (spur) between r and $r + dr$ is given by the Poisson law:

$$dp(r|spur) = 2\lambda r dr. \quad (8)$$

We adopt the local surface density of sources at least as bright as m_0 , the magnitude of the candidate. Because more sources are available in a same given area, the densities computed with this method are more local – or more accurate – than densities computed in arbitrary bins of magnitudes. It is equivalent to computing local densities using increasingly sensitive instruments. We detail in [Appendix B](#) the method used to estimate local densities.

The likelihood ratio is the ratio of the two probability densities (7) and (8):

$$LR(r) = \frac{dp(r|cp)}{dp(r|spur)} = \frac{1}{2\lambda} e^{-\frac{1}{2}r^2}. \quad (9)$$

The formalism we apply here aims at providing probabilities of identification based on positional coincidences only. A Bayesian interpretation of the likelihood ratio method is described in [Appendix C](#). We do not use other information on sources such as the spectral energy distribution. Hence, we do not add an extra term $q(m)$ to the LR as is done for example in [Wolstencroft et al. \(1986\)](#), [Sutherland & Saunders \(1992\)](#) and [Brusa et al. \(2007\)](#).

The quantity $q(m)$ corresponds to the probability of having among the real counterparts a source of magnitude m , or in a bin Δm around m (see formula (C.8) of the Appendix). In this case, $q(m)$ should be local, but then becomes hard to estimate. In general the estimate of $q(m)$ is plagued with considerable errors which, besides the error on the local density estimation, dramatically affect the error on LR . We will see in [Sect. 3.3](#) that the $q(m)$ factor is somehow taken into account in our reliability function.

3.3. Computing reliabilities

Although we use a different LR definition, a different estimator of the rate of spurious associations and a different function to fit the reliability histogram, we more or less follow the work presented in part 3 of [Oyabu et al. \(2005\)](#). The method originates in [Rutledge et al. \(2000\)](#).

We define the reliability of an association in a given bin of LR as

$$R(LR) = \frac{N_{\text{real}}(LR)}{N_{\text{real}}(LR) + N_{\text{spur}}(LR)} = \frac{N_{\text{cand}}(LR) - N_{\text{spur}}(LR)}{N_{\text{cand}}(LR)}, \quad (10)$$

where N_{real} and N_{spur} are the unknown number of candidates which are respectively real and spurious counterparts in a given bin of LR ; N_{cand} is the number of candidates in a given bin of LR .

We therefore have to estimate N_{spur} . An often used method consists in correlating X-ray sources with artificial samples of optical sources. The generated samples have the same characteristics as the real sources: same density, same positional errors distribution, etc. Positions are randomly distributed. The sum of the results of these Monte-Carlo samples provides an estimate of the number of spurious associations as function of the distances, of the LR s, etc. This approach is used by [Oyabu et al. \(2005\)](#) among others. In [Stephen et al. \(2005\)](#) the random sample consists in a list of “anti [...] sources”, which are “mirrored in Galactic longitude and latitude”.

We propose here a new method to estimate the number of spurious associations, not based on Monte-Carlo simulations, but instead directly computing their expected results. This scheme offers a better computing efficiency when cross-correlating huge sets of data. The basic idea of estimating the surface of an association related to the total available area can be found in [Boller et al. \(1998\)](#). The method is described in [Appendix D](#).

In order to avoid computing too many local densities for estimating the rate of spurious associations, we divide the magnitude range into bins and associate all sources in the same magnitude bin with the mean value of their local density. The width of the bins depends on the magnitude accuracy of the catalogue. We then compute for all optical and X-ray sources the $\sigma_M\sigma_m$ factor, the LR_{min} and LR_{max} . It is thus possible to compute the histogram of the expected number of spurious associations according to LR values. To increase the computing efficiency, we can bin the $\sigma_M\sigma_m$ values. However, this approach involves another loss of accuracy for a meagre reduction of computing time.

As shown in [Fig. 2](#), the histograms used in the computation of the reliability are the number of candidates and the number of spurious associations grouped in bin of $\log_{10} LR$.

3.3.1. Fitting the reliability function

In order to estimate the number of spurious associations we take the relatively realistic example where the X-ray source has at

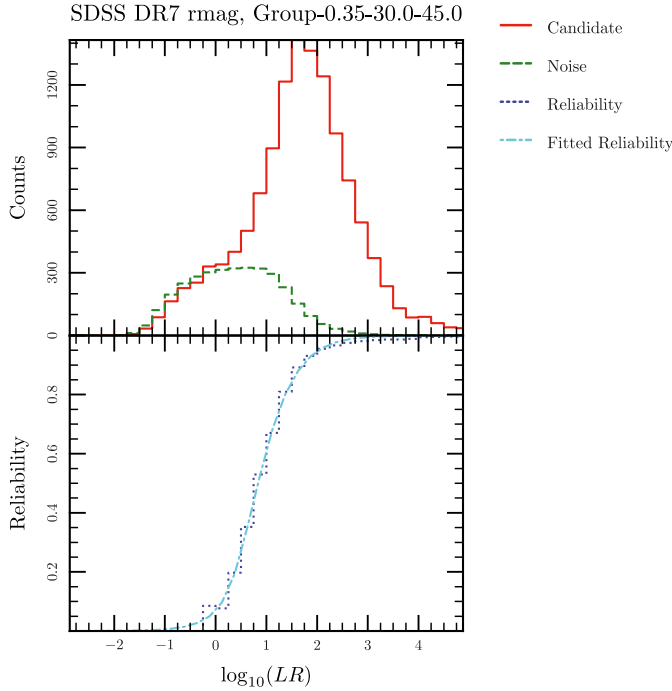


Fig. 2. *Top:* histograms of the number of associations and of the estimate of the number of spurious associations by bin of $\log_{10}(LR)$. *Bottom:* Reliability histogram by bin of $\log_{10}(LR)$ and its fitted curve. In the example, LR s have been computed according to the SDSS DR7 r magnitude for XMM sources with a systematic error of 0.35 arcsec with a galactic latitude $30^\circ < |b| < 45^\circ$.

most one candidate. The reliability of an association (not to be confused with the integrated reliability for all associations having a $LR \geq l$) is directly given by (see Eq. (C.6) of the Appendix)

$$R(LR) = p(id|r) = \frac{1}{1 + \frac{p(spur)}{p(cp)} \frac{1}{LR}}. \quad (11)$$

The term $p(cp)/p(spur)$ – the probability that the optical source is a counterpart divided by the probability that it is spurious – must be independent of the dimensionless distance r (see Eq. (6)). It is similar to the term $(1 - \theta)/\theta$ used in De Ruiter et al. (1977). However, $p(cp)/p(spur)$ may depend on the nature of the underlying X-ray source population (e.g. stars, AGN) and may thus vary with source properties such as magnitude, optical colour, or flux ratios. In order to obtain a LR similar to that used in Brusa et al. (2007) for instance, we would need to consider an additional parameter $q(m)$ describing the variation of $p(cp)/p(spur)$ with the magnitude (or any other relevant property) of the candidate counterpart. Alternatively, $q(m)$ may be replaced by another term such as that playing the role of B_{phot} in Budavári & Szalay (2008). An $R(LR)$ histogram can be built from the $N_{\text{cand}}(LR)$ and $N_{\text{spur}}(LR)$ histograms made using the method explained in the previous paragraph. If the ratio $p(cp)/p(spur)$ were independent of source properties, $R(LR)$ could be fitted with Eq. (11) using only one free parameter $a = p(cp)/p(spur)$. Including a term $q(m)$ in LR with $N_{\Delta m}$ bins of magnitude, requires us to build $N_{\Delta m}$ $R(LR, \Delta m)$ histograms and fit each of them with functions in Eq. (11) having different a parameters. However, in general the lack of statistics does not allow us to do so.

The $R(LR)$ histogram can then be seen as the sum of $N_{\Delta m}$ $R(LR, \Delta m)$ histograms and consequently can be modelled

by the function

$$R_f(LR) = \sum_{i=1}^{N_{\Delta m}} \frac{b_i}{1 + a_i \frac{1}{LR}}, \quad \text{with } b_i = \frac{N_{\text{cand}, \Delta m_i}}{\sum_{j=1}^{N_{\Delta m}} N_{\text{cand}, \Delta m_j}}, \quad (12)$$

where $N_{\text{cand}, \Delta m_i}$ is the total number of entries in histogram number i .

In practice, the histograms are not binned according to LR but to $\log_{10}(LR)$. Best fits were obtained using the function

$$R_f(x = \log_{10}(LR)) = \frac{\sum_{i=1}^{N_{\Delta m}} b'_i 10^{(1-i)x}}{1 + \sum_{i=1}^{N_{\Delta m}} a'_i 10^{-ix}} \quad (13)$$

with $N_{\Delta m} = 3$, i.e. 6 free parameters.

The fit is performed using a Levenberg-Marquard algorithm. We compute the same number of LR and construct and fit the same number of LR histograms as there are magnitude bands in the SDSS.

3.4. Computing probabilities of identification in the general case

We now extend the Bayesian approach to X-ray sources having N_{cand} candidates. We assume that at most one association is real. This assumption should be fulfilled in our case for at least two reasons. First, we only consider point-like X-ray sources. This condition decreases the probability that the detection results from two distinct unresolved sources blended in the XMM-Newton beam. Second, 95% of the XMM-Newton sources matching a SDSS entry with a probability higher than 90% have a 0.5–2.0 keV flux higher than 1.65×10^{-15} erg cm $^{-2}$ s $^{-1}$. At this flux, source confusion is of the order of a few percent only (Cappelluti et al. 2009). The corresponding source density of ~ 500 deg $^{-2}$ (Cappelluti et al. 2009) is well below the value of 2000 deg $^{-2}$, above which simulations show that source confusion becomes important (Loaring et al. 2005). Similar conclusions can be drawn for the hard (2–12 keV) sources.

Let us consider $N_{\text{cand}} + 1$ hypotheses:

- H_{cp_i} : the i th optical source is the counterpart;
- $H_{spur_{\text{all}}}$: there is no counterpart.

Then the Bayesian probability that the i th source is the counterpart knowing r_j , $j \in [1, N_{\text{cand}}]$ is

$$P_{id,i} = R'_i = p(H_{cp_i} | r_1 \cap \dots \cap r_{N_{\text{cand}}}) = \frac{LR_i \frac{p(H_{cp_i})}{p(H_{spur_{\text{all}}})}}{1 + \sum_{j=1}^{N_{\text{cand}}} LR_j \frac{p(H_{cp_j})}{p(H_{spur_{\text{all}}})}}. \quad (14)$$

If $\frac{p(H_{cp_i})}{p(H_{spur_{\text{all}}})} = \frac{p(cp_i)}{p(spur_i)}$, Eq. (14) leads to the formula below, obtained following the Rutledge et al. (2000) prescription

$$P_{id,i} = R'_i = \frac{\frac{R_i}{1-R_i}}{1 + \sum_{i=1}^M \frac{R_i}{1-R_i}}, \quad (15)$$

with Eq. (11), we easily show that $R/(1-R) = LRp(cp)/p(spur)$. Computing $p(H_{cp_i})$ and $p(H_{spur_{\text{all}}})$ normalising the terms $p(cp_i)$, $p(spur_i)$ as Rutledge et al. (2000) do to construct $P_{id,i}$ from R_i ,

Table 2. Number of unique 2XMMi sources (N_x) with N_c SDSS DR7 candidates.

N_c	1	2	3	4	5	6	7	8	>8
N_x	37 988	6059	1196	317	96	47	12	4	8

we obtain the equality $\frac{p(H_{cp_i})}{p(H_{spur_{all}})} = \frac{p(cp_i)}{p(spur_i)}$. We thus apply Eq. (15) to compute the final probabilities of identification.

Each candidate possesses as many reliabilities as there are magnitude bands in the SDSS. The R_i we consider in the final probability of identification formula are for each source the best of all photometric bands.

4. Observation grouping

XMM-Newton EPIC sources are correlated FOV by FOV, ie, observation by observation. In order to tail off count-rate noise on FOV LR histogram bins without sacrificing resolution, we have to increase count statistics. We therefore stacked data from similar FOV:

- we split into two groups XMM-Newton FOV with different systematic errors on position: 0.35'' or 1.0'';
- observations of the LMC and SMC regions are set apart;
- because they presumably share objects of same nature and same patterns of $\log N$ - $\log S$ relation, observations are grouped according to their galactic latitude.

As mentioned above, the relation between reliability and likelihood ratio depends on the overall properties of the X-ray populations present in the optical sample. In addition to galactic latitude, we also tested whether the X-ray flux could significantly modify the shape of the $R(LR)$ curves. Splitting further 2XMMi sources into groups of medium (10^{-14} – 10^{-13} erg cm $^{-2}$ s $^{-1}$) and faint (10^{-15} – 10^{-14} erg cm $^{-2}$ s $^{-1}$) 0.2–12 keV flux ranges does not change the probabilities of identification by more than ~3% in most cases. The only noticeable difference is for faint sources with identification probabilities below 50%, which tend to show even lower identification probabilities by as much as 15%. We felt, however, that since the effect is relatively modest and only affects sources for which the significance of the identification is rather low, priority should be given to the gathering of sufficiently large subsamples. We thus did not consider any X-ray flux dependency in the final implementation.

5. Results of the 2XMMi-SDSS DR7 cross-correlation

A total of 1337 XMM-Newton FOV hold at least one source with a SDSS counterpart candidate within the combined 3σ search radius. These 1337 FOV contain 95 452 detections, corresponding to 73 636 unique 2XMMi sources. The cross-correlation of the 2XMMi catalogue with the SDSS DR7 leads to 72 169 “associations” involving 45 727 and 55 726 unique 2XMMi and SDSS DR7 sources respectively. This first number represents 20% and 62% of the unique sources available in the entire 2XMMi catalogue and in the 1337 FOV respectively. The distribution of the number of SDSS DR7 candidates by unique 2XMMi sources is given in Table 2, and the main properties of the distribution of the probabilities of identification and of their cumulative values are given in Table 3. We define the sample completeness as the fraction of 2XMMi sources having an individual probability of identification in SDSS above a given cutoff relative to the total

Table 3. Cross-correlation statistics.

id	Probability of identification cutoffs					
	0.0	0.5	0.7	0.8	0.9	0.95
# det_id	60 567	53 347	49 527	46 387	40 193	32 610
# src_id	45 727	39 839	36 943	34 605	30 055	24 327
R	79.4%	91.7%	94.7%	96.2%	98.0%	99.0%
C	100.0%	96.8%	92.1%	87.5%	77.2%	63.2%
Frac X	62.1%	54.1%	50.2%	47.0%	40.8%	33.0%

Notes. Number of associations with all XMM detection (det_id) and all unique sources (src_id) with a probability of identification above a given probability of identification. If an X-ray source has several candidates, we only keep the one with the best probability of identification. Sample reliability (R) and sample completeness (C): we only consider the best match for each unique XMM sources having at least one counterpart. Frac X is the expected fraction of X-ray sources with a counterpart in the SDSS DR7 catalogue.

number of 2XMMi sources with SDSS counterparts. In a similar manner, sample reliability is the fraction of non-spurious associations among 2XMMi sources having an individual probability of SDSS association above a given threshold. A total of 7740 unique 2XMMi sources have several SDSS DR7 candidates. In this sub-sample, there are 896 and 2 672 2XMMi sources for which the candidate with the highest identification probability is not the nearest and the brightest SDSS DR7 candidate respectively.

The left panel of Fig. 3 shows the distribution of the individual SDSS source identification probabilities. Most SDSS entries found within the combined 3σ search radius from the 2XMMi source have a high likelihood to be the true optical counterpart. The small tail of very low identification probability objects reflects the expected rising contribution of SDSS entries unrelated to the X-ray source at large matching distances. Most SDSS entries with identification probability higher than ~90% are found less than 3 arcsec from the X-ray position (Fig. 3, centre and right panel). The rather wide spread of the 2XMMi positional errors accounts for the scatter affecting the distances at which high-probability SDSS sources are found from the X-ray position.

Expressed in terms of combined 2XMMi + SDSS errors, the distance distribution shown in the right panel of Fig. 3 follows the usual shape of a Rayleigh distribution. Fitting this histogram with a Rayleigh function plus a linear component, we obtain $\sigma_r = 0.865$ for the Rayleigh curve parameter, 0.178 for the linear slope and $R = 0.675$ for the ratio between the total number of real associations and the total number of spurious ones within the search radius. Fitting separately distance histograms of the sources whose positions were corrected by *eposcorr* and uncorrected ones leads to $\sigma_r = 0.856$, $R = 0.698$ and $\sigma_r = 1.013$, $R = 0.50$ respectively. All errors on σ_r , on the slope and on the R values are about 0.003. Keeping only the best candidate for each unique XMM source, we obtain $R=0.81$, which is consistent with the value of 79.4% given in Table 3. The origin of this small apparent overestimate (~14%) of the positional errors of *eposcorr* corrected sources is so far unclear. In any case, the effect of this slight change on the identification probabilities is small. The global effect is to slightly decrease the probabilities of SDSS entries matching at large distances and to somewhat increase the probabilities of those located close to the X-ray source.

The practical implementation is described in Appendix E. Whenever optical data are used, we discard SDSS entries with recorded magnitudes fainter than 22.2 in any of the

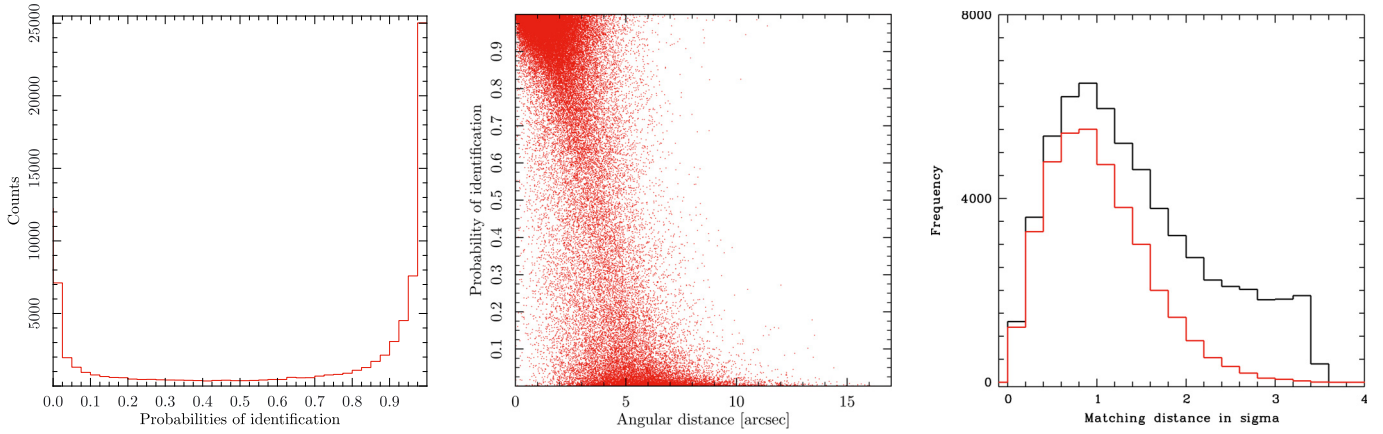


Fig. 3. *Left:* histogram of the individual probabilities of identification. *Centre:* probabilities of identification versus matching distances. *Right:* distribution of the distance of SDSS candidates to 2XMMi sources expressed in units of the combined 2XMMi + SDSS positional error; black = all matches, red = identification probabilities $\geq 90\%$.

photometric bands considered. Indeed, objects with magnitudes higher than 22.2 tend to have smaller photometric errors than brighter ones, clearly indicating that SDSS photometric uncertainties and perhaps also mean values are not reliable at faint flux. We also ignored all SDSS entries having one of the following flag set: `BLENDED`, `DEBLENDED_AS_MOVING`, `SATURATED`, `INTERP_CENTER`, `EDGE`, `SATUR_CENTER`, `PSF_FLUX_INTERP` in order to ensure the best photometric quality. Unless specified otherwise, we will hereinafter only consider optical identifications with a probability larger than 90%. This threshold applies to both the spectroscopically identified sample and to the general photometric sample and corresponds to an overall sample purity of 98% (see Table 3).

6. Building an identified sample

One of the important task given to the SSC is the statistical identification and classification of all X-ray sources discovered in the wide field of view of the EPIC cameras. The statistical determination of the nature of any given 2XMMi source will first rely on the assessment of the reliability of its association with candidate counterparts at other wavelength. The description of this important step and of its results are the goals of the present paper.

On the other hand, the subsequent classification stage requires the knowledge of the parameter space occupied by the various groups of astrophysical sources using a “learning sample”. Therefore, the cross-correlation method presented here allows us to select in a clean and statistically controlled manner the best optical counterparts to 2XMMi sources and constitutes the first mandatory step towards building a reliable learning sample, which can be later used to define source classification schemes using advanced statistical methods. Eventually, the classification method, either supervised or not, will provide the most likely nature of the 2XMMi source (e.g. star, AGN, etc.) with for some methods, an estimate of the probability of the classification. First attempts to classify 2XMMi sources in two classes (stars and extragalactic) have been presented in Pineau et al. (2009) and are now implemented in the XCat-DB for the DR3 of the 2XMM catalogue.

Yet it is also well known that a reliable classification can only be achieved when the corresponding learning sample

covers the parameter space spanned by the group of objects to identify as evenly as possible, see e.g. White (2008) or Richards et al. (2004). Being aware of this important requirement, the SSC has designed a general optical identification programme able to explore the widely diverse natures of the X-ray emitting objects discovered in the XMM source catalogues. Several wide field identification campaigns are currently conducted at various X-ray flux levels and galactic latitudes, which all aim at building completely identified source samples. The nature of the high b population is the scope of four distinct projects. The bright part is studied by the Bright Sources Survey (XBS or BSS, Della Ceca et al. 2004; Caccianiga et al. 2008). The XMM-Newton Medium Sensitivity Survey (XMS, Barcons et al. 2002; Carrera et al. 2007; Barcons et al. 2007) and the XMM-2dF Wide Angle Survey (XWAS, Tedds et al. 2006) investigate the properties of medium flux sources. The faintest source population is the scope of the Subaru/XMM-Newton Deep Survey (SXDS, Ueda et al. 2008). Finally, the Galactic plane area is covered by the XMM-SSC Galactic Plane Survey (Motch 2005; Motch et al. 2010).

The first step towards building a sample of 2XMMi sources of known astrophysical nature was to select X-ray sources with reliable SDSS DR7 spectroscopic counterparts of a known class (i.e., with the *specClass* attribute pointing to an astrophysical object). For our purpose, the three most important groups of spectroscopic SDSS targets are the sample of quasar candidates defined by Richards et al. (2004), the main galaxy sample described in Strauss et al. (2002) and all stars belonging to the legacy survey and to the Sloan Extension for Galactic Understanding and Exploration programme (SEGUE, Yanny et al. 2009). The AGN sample is mostly a classically UV-excess (UVX) selected sample to which is added a small number of redder targets appearing as likely high redshift QSOs. The galaxy sample is less biased because it is only selected on brightness related criteria in the r band. Stars from the legacy survey were mostly selected on the basis of their extreme colours. Among them, red dwarfs and CVs are the most likely to match 2XMMi sources. The SEGUE programme opens new areas at lower galactic latitudes, and its spectroscopic target selection aims at covering all spectral types.

We therefore extracted the SDSS spectroscopic catalogue accessible via CasJob, and following the SDSS spectral class

scheme, define the classes: stars, galaxies, AGN and X-ray accreting binaries. We list below the origin of the different groups of identified sources:

- Stars : i) 2XMMi/SDSS associations having the *specClass* attribute set to 1 or 6 and ii) the sample of stars coming from the kernel density classification (see Sect. 6.1).
- Accreting binaries : i) objects in the Downes catalogue of cataclysmic variables (Downes et al. 2001). We used here the 2006 version, which contains many SDSS discoveries; ii) the Ritter catalogue of cataclysmic variables (Ritter & Kolb 2003) and iii) the Ritter catalogue of LMXRBs (Ritter & Kolb 2003).
- Galaxies : 2XMMi/SDSS associations with a probability of identification >0.80 and with the *specClass* attribute set to 2.
- AGN: i) sources from the Véron catalogue (Véron-Cetty & Véron 2006) and ii) SDSS DR7 objects associated with a 2XMMi source with a probability of identification >0.80 and having the *specClass* attribute set to 3 or 4 (QSO or high z QSOs).

We use the range of X-ray luminosity to define several groups of active galaxies and consider all extragalactic objects as a single class. In particular, we do not make any formal distinction between QSO and AGN². An X-ray source associated with both a star and an accreting binary was flagged as an accreting binary. We applied the same rules for star-AGN and binary-AGN pairs of apparently conflicting nature.

We added sources identified in the XBS and XMS SSC surveys and with SDSS counterparts. For the XMS, we considered their sources with classes NELG, BLAGN, and BLLac as AGN. For the XBS, AGN2, AGN1, BLLac and elusive AGN were assigned the general AGN type.

QSO2s candidates taken from Zakamska et al. (2003) and Reyes et al. (2008) as well as a handful of X-ray selected objects (see Sect. 8.1 below) having a reliable match in the 2XMMi catalogue were added to the identified sample.

6.1. The stellar identified sample

Building a clean stellar sample turned out to be more difficult because most stellar sources detected in X-rays have optical SDSS magnitudes brighter than 15 mag and are flagged as saturated. Furthermore, the SDSS DR7 spectroscopic database provided only few cross-matches with acceptable properties (i.e. non-saturated and probabilities of identification higher than 90%). Therefore, in order to enlarge the stellar sample, we applied a classification method allowing us to identify stars on the basis of their multi-colour properties.

We performed a kernel density classification (KDC, Richards et al. 2004) on all spatially unresolved ($cl=6$) SDSS candidates. This selection returns 10533 SDSS sources with a correlation in the 2XMMi catalogue. The classification only uses the four colours $u - g$, $g - r$, $r - i$ and $i - z$ as parameters. The learning sample used for this classification consists of two classes: star and QSO, since we only consider point-like objects in the optical. It has been built from all unresolved SDSS sources, independently of their association with a 2XMMi entry. We only retained good

² A large fraction of these AGN are found the Véron catalogue (Véron-Cetty & Véron 2006), which is based on the DR4. Among the 1290 SDSS DR7 spectroscopic QSOs entries, 836 are also present in the Véron catalogue and are classified QSOs or AGN according to their absolute blue magnitude MB.

Table 4. Results of the classification method applied to the learning sample.

org\assign	Star	AGN
Star	96.89%	3.11%
AGN	1.62%	98.38%

quality detections (i.e. no flag SATURATED, BLENDED, DEBLENDED_AS_MOVING, INTERP_CENTER, EDGE, SATUR_CENTER or PSF_FLUX_INTERP set) that were spectroscopically identified in the DR7. The data have been retrieved from the DR7 database with CasJob. The stellar sample contains 67269 sources flagged by the SDSS *specClass* attribute as star (STAR or STAR_LATE) and therefore also contains CVs and WDs. The non-stellar sample has 75248 sources flagged by the SDSS *specClass* attribute as QSO (QSO) or high-redshift QSO (HIZ_QSO) plus 253 sources flagged by *specClass* as galaxy (GALAXY). For simplicity we call the non-star sample AGN sample below.

Estimates of the probability densities were computed using a fixed bandwidth kernel smoothing. The kernel applied uses the Epanechnikov profile and the bandwidth was chosen to be equal to 0.2 mag. Table 4 lists the results of the self-check of the learning sample, i.e., the results of the classification method applied to the learning sample only.

The prior probability $p(\text{star})$ has been set to 0.25 and so $p(\text{AGN})$ to 0.75 as a result of iterative kernel density classifications converging to this relative number of stars and AGN in the SDSS/2XMMi learning sample. In order to select SDSS/2XMMi identifications with the best chance to be normal stars, we removed 13% of all SDSS entries classified as stars, but falling in low-density regions of the parameter space (i.e. far from the centre of the stellar multi-colour locus) and thus prone to be doubtful cases such as binaries, unidentified cataclysmic variables, or even mis-identified AGN.

The star/AGN classification has been made according to the optical properties of the spectroscopically identified SDSS objects. However, by construction, the density distribution in colours of the SDSS/2XMMi sample is not likely to follow that of the non X-ray emitting SDSS objects and can thus lead to some biases. For instance, there may be a considerable over-density of non-X-ray emitting stars in some part of the 4-d colour diagram where most objects classified as AGN appear to be strong X-ray sources. This problem indeed occurs in the region covered by the AGN branch, where there is some overlap with A stars. Although some A stars do emit X-rays for debated reasons, not all do. We thus removed from the SDSS/2XMMi learning sample all classified stars with a $u - g$ colour of less than 1.2 (values taken from Covey et al. 2007).

The final stellar X-ray sample arising from the KDC contains 636 unique entries with a classification probability higher than 99.7% (3 Gaussian σ). However, only 549 of these matches have a probability of identification with a 2XMMi source higher than 90% and were therefore entered in the final identified sample.

We also checked that the stellar SDSS/2XMMi sample adhered to the stellar locus derived by (Covey et al. 2007) using synthetic photometry in the 4-D colour space. The agreement is good, apart for the reddest stars of spectral type later than $\sim M5$.

6.2. The final identified sample

The origin and distribution of the various classes of identified objects in the final sample are listed in Tables 5–7. Most of the extragalactic identified sample comes from the DR7 spectroscopic

Table 5. Distribution of XMS and XBS sources in the identified samples.

Sample	Galaxy		AGN		QSO	
	XMS	XBS	XMS	XBS	XMS	XBS
All	1	0	7	21	49	95
Final	1	0	5	14	37	85

Notes. The “Final” sample only contains associations with a probability of identification >0.9 , with SDSS sources unsaturated, not blended, ..., and where all magnitudes <22.2 mag.

Table 6. Origins of the objects of each type in the final identified sample.

Sample	Star		Accreting Binary			Extragalactic	
	DR7	KDC	Dow. ^a	R1 ^b	R2 ^c	DR7	Veron
Final ^d	8	541	29	22	2	2021	1524

Notes. ^(a) Downes catalogue of CV. ^(b) Ritter catalogue of CV. ^(c) Ritter catalogue of LMXB. ^(d) Note that individual sources can be listed in several catalogues.

Table 7. Distribution of number of unique 2XMMi entries with types in the final identified sample.

Sample	Star	Accr. Binary	Extragalactic
Sub	549	26	2336

catalogue through the Véron catalogue, while the vast majority of X-ray active stars are actually extracted from the KDC source classification. A large fraction of the identified accreting binaries (mostly cataclysmic variables) also come from follow-up SDSS discoveries through the Downes catalogue.

Finally, for all AGNs with a spectroscopic redshift we compute the observed X-ray luminosity using

$$L_{\text{abs}} = 4\pi F \frac{(1+z)^2}{(1+z)^{2-\gamma}} \left(\frac{c}{H_0} \int_0^z \frac{1}{\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}} dz \right)^2, \quad (16)$$

where F is the 0.2–12 keV X-ray flux in $\text{erg s}^{-1} \text{cm}^{-2}$, $H_0 = 73 \text{ km s}^{-1} \text{Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. The photon index γ was taken to be 1.9.

7. Grouping X-ray sources in parameter space

7.1. Sample properties and shortcomings

The left and centre panels of Fig. 4 display the X-ray flux distributions of 2XMMi sources with individual identification probabilities with SDSS DR7 entries $\geq 0\%$ and $\geq 90\%$, compared to that of all 2XMMi sources present in the SDSS DR7 footprint. Many faint X-ray sources have likely counterparts in the SDSS. Figure 4 also shows that the fraction of 2XMMi sources with likely SDSS identifications does not vary strongly with X-ray flux. It steadily decreases by a factor of 2 from $F_X \sim 1 \times 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$ to $1 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$. The drop in the identification rate at a flux above $\sim 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$ is probably caused by the increasing number of bright optical counterparts, which are likely to be flagged as saturated in SDSS and therefore absent from our sample. On the other hand, the shape of the decline of the SDSS identified fraction with decreasing X-ray flux (centre panel) as well as the observed distribution of the f_x/f_{opt} ratio with X-ray flux (right panel) are both consistent

with populations of X-ray sources with a weakly varying distribution of f_x/f_{opt} ratios with X-ray flux. In other words, comparing the SDSS identified sample with the total sample does not reveal evidence of strong evolution of f_x/f_{opt} with redshift. A comparable conclusion was reached in Sect. 4 based on the weak dependency with X-ray flux of the reliability / likelihood ratio relations. It may thus be possible to extrapolate the properties of the 2XMMi/SDSS DR7 photometric identified sources to somewhat fainter X-ray and optical fluxes.

The situation of the “final” spectroscopic identified sample clearly differs. Its X-ray flux distribution strongly differs from that of the 2XMMi/SDSS DR7 photometric sample and from that of the overall 2XMMi sample. Obviously, this discrepancy arises from the higher optical brightness needed by spectroscopic observations (see left panel of Fig. 4). In addition, the choice of the spectroscopic targets results from various heterogeneous optical selection criteria and is therefore unlikely to cover all X-ray emitting objects above equally any given X-ray flux threshold. Some examples are outlined in the sections below. We therefore stress that as it stands, this identified sample cannot in any manner be used as a learning sample suitable for a statistically reliable classification of 2XMMi sources with SDSS identifications.

However, the high number of spectroscopic SDSS matches supplemented by other identifications derived from archival catalogues allows us to build an unprecedentedly large sample of X-ray sources of known nature, enriched with accurate multi-colour photometry and detailed spectral line measurements. This large collection of best quality data offers a unique opportunity to study to some extent the parameter loci occupied by the different classes of X-ray emitters. However, it also allows addressing two important issues. First, finding the most efficient physical parameters for separating different groups of X-ray sources. Second, it allows us to highlight the parameter regions not well covered by the SDSS observing strategy and therefore in need of extended spectroscopic studies.

The huge merit of the dedicated identification programmes such as the ones carried out by the SSC is to extend spectroscopic identifications to very low optical fluxes and thus offer a unique opportunity to unveil the different populations of extragalactic sources which may appear at fainter optical fluxes and in general at higher X-ray to optical flux ratios. The two strategies, wide and shallow on one hand and narrow and deep on the other hand are indeed quite complementary and suited to best characterise and scientifically investigate the entire serendipitous XMM-Newton catalogues.

7.2. The main classes of X-ray sources

We investigate the distribution of the various classes of X-ray sources in the original instrumental parameter space. In principle, we could also have used the parameter space resulting from a Principal Component Analysis (PCA), thus highlighting the most significant (or information-rich) linear combination of physical measurements. Pineau et al. (2008) showed that the two first eigenvectors deriving from the PCA analysis of the 2XMMi/SDSS DR7 sample gathering the largest data variance are indeed close to the two main ones used here, namely X-ray to optical flux ratios and optical colours. However, taking into account eigen-axes of higher orders, which include X-ray spectral information in the form of hardness ratios, can slightly improve the separation between different classes of sources (Pineau et al. 2008).

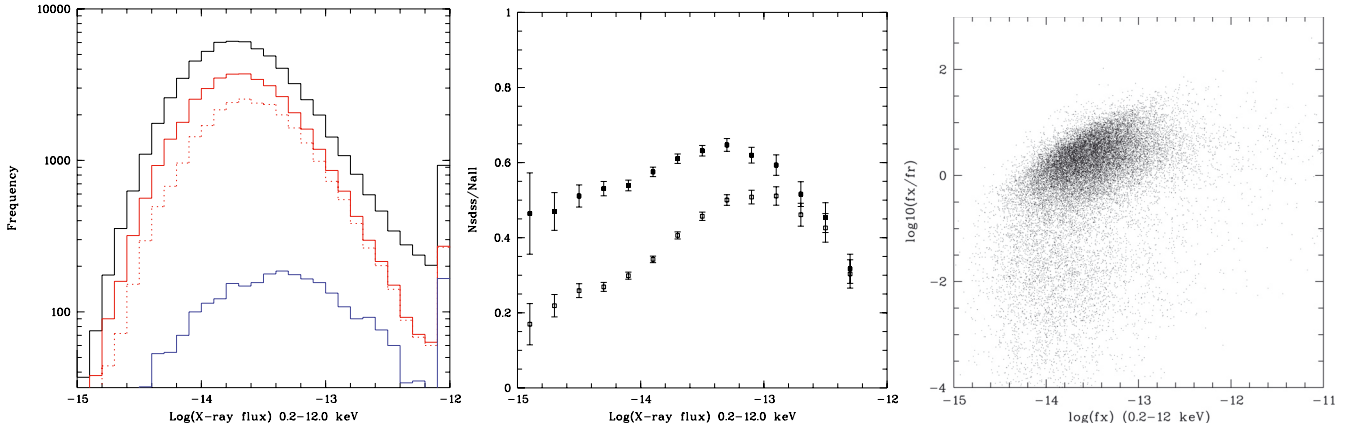


Fig. 4. *Left panel:* X-ray flux distribution. *From top to bottom,* black line: all X-ray sources found in 2XMMi fields overlapping SDSS DR7, red line: all X-ray sources matching a SDSS entry, red dashed line: 2XMMi/SDSS R7 correlations with individual identification probabilities $>90\%$, blue line: “final” identified sample. *Centre panel:* ratio of the SDSS identified to total number of 2XMMi sources. Upper curve, filled squares; all matches, lower curve, empty squares, matches with identification probabilities $\geq 90\%$. *Right panel:* variation of $\log(f_x/f_r)$ with X-ray flux for 2XMMi/SDSS DR7 sources with identification probabilities $\geq 90\%$ and r magnitude brighter than 22.2. In all cases, the combined unique EPIC source detection likelihood is ≥ 6 .

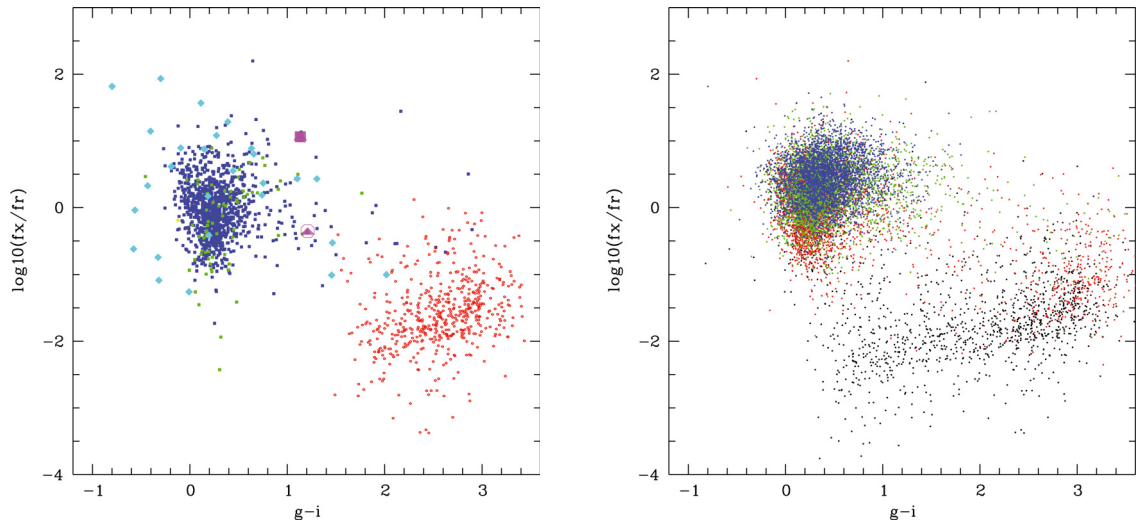


Fig. 5. Distribution of spatially unresolved objects in the optical band in the $\log(f_x/f_r)$ versus $g-i$ diagram. *Left:* the identified sample. Blue: AGNs with $\log(L_X) \geq 44$, green: $44 \geq \log(L_X) \geq 42$, magenta: QSOs – filled squares = X-ray selected – filled triangles = optically selected – encircled = Compton Thick (see Sect. 8.1), red: stars, cyan: accreting binaries. *Right:* the entire SDSS photometric sample. In this case, the colour codes the range of r magnitude. black: <18 , red: 18–20, green, 20–21, blue: >21 . We only show SDSS entries with a probability of identification with an X-ray source higher than 90%, g and i magnitudes brighter than 22.2 and errors on $g-i < 0.2$.

We show in Figs. 5 and 6 the positions in the $g-i / \log(f_x/f_r)$ diagram of the various classes of objects present in the identified sample (left panel) and of all 2XMMi sources with only a cross-identification with a photometric SDSS entry (right panel). Sources spatially unresolved and extended in the optical band are presented separately in the two figures. A majority of 2XMMi sources match with SDSS-DR7 photometric entries close to the limiting magnitude of the SDSS survey (mag ~ 22), and only a relatively small fraction is bright enough to have been selected for spectroscopic observations. For instance, over a grand total of 60567 2XMMi detections matching a SDSS DR7 entry, 87% have an error on their r magnitude below 0.2, but only 12% are SDSS spectroscopic targets as well. In addition, the repartition of the spectroscopic targets are far from covering uniformly the parameter space spun by the optical counterparts of the serendipitous XMM-Newton sources.

7.3. Separating stellar from extragalactic sources

As expected, the f_x/f_r ratio is a very powerful parameter to separate the late-type stellar X-ray population in which the

high-energy emission arises in a magnetic active corona from X-ray luminous sources powered by accretion such as active galactic nuclei or cataclysmic variables. However, the distribution of low L_X galaxies in $\log(f_x/f_r)$ clearly overlaps with that of active coronae. Introducing the $g-i$ colour index allows us to separate the bulk of the stars, especially the reddest M stars from most galaxies. Nevertheless, many galaxies, in particular of the early type, exhibit optical energy distributions similar to those of G type stars, show comparable (f_x/f_r) and consequently cannot be easily distinguished from stars in the $g-i / \log(f_x/f_r)$ diagram. Obviously, taking into account the spatial extension of the optical source allows us to efficiently separate them from stars (see Figs. 5 and 6).

Interestingly, the reddest point-like optical sources located on the “stellar” branch are also the faintest ones with r magnitudes in the range of 18 to 20 (see Fig. 5, right panel). They also appear to exhibit the highest f_x/f_r ratio. This is consistent with the known increase of the f_x/f_{opt} ratio for M stars compared to that of earlier spectral types (see e.g. Vaiana et al. 1981). We note, however, that some high z QSOs have been identified with

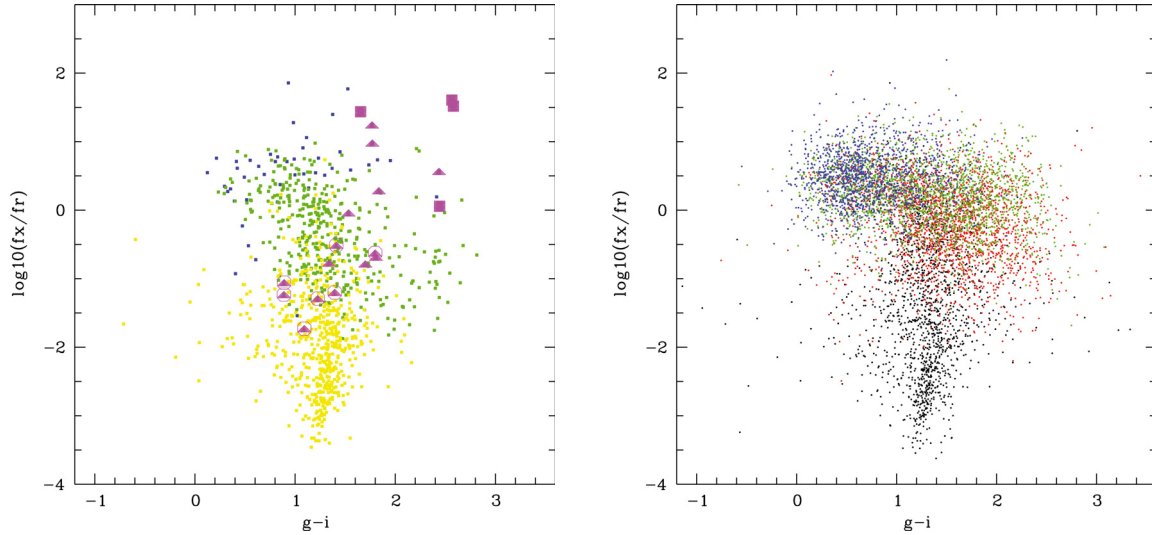


Fig. 6. Distribution of spatially resolved objects in the optical band in the $\log(f_x/f_r)$ versus $g-i$ diagram. *Left*: the identified sample blue: AGNs with $\log(L_X) \geq 44$, green: $44 \geq \log(L_X) \geq 42$, yellow: galaxies with $\log(L_X) \leq 42$ AGN, magenta: QSO2 – filled squares = X-ray selected – filled triangles = optically selected – encircled = Compton Thick (see Sect. 8.1), red: stars, cyan: accreting binaries. *Right*: entire SDSS photometric sample. In this case, the colour codes the range of r magnitude. Black: <18 , red: $18-20$, green, $20-21$, blue: >21 . We only show SDSS entries with a probability of identification with an X-ray source larger than 90%, g and i magnitudes brighter than 22.2 and errors on $g-i < 0.2$.

very red point-like objects of f_x/f_r ratios approaching those of active coronae (see Fig. 5).

Although cataclysmic variables occupy a locus in the $g-i / \log(f_x/f_r)$ comparable to that of most quasars, their distribution exhibits a wider spread than that of AGNs. This large scatter can be used to provide a high likelihood identification of their class, at least for part of them. For instance, very blue objects, typically with $g-i$ below -0.2 as well as those with extreme f_x/f_r , have a high probability of being cataclysmic variables.

7.4. Distinguishing between the various classes of extragalactic sources

Figure 5 shows that many of the 2XMMi sources having a counterpart in the DR7 of the SDSS cluster in a rather narrow range of blueish $g-i$ colours in the interval of -0.2 to 0.8 . They are characterised by a $\log(f_x/f_r) \sim 0$ and appear as point-like sources in the optical. Their positions in this diagram overlaps with that of the vast majority of the spectroscopic SDSS AGN found in our identified sample, which for most of them are UV-excess optically selected quasars.

Let us now consider all objects, both spatially resolved and unresolved, occupying the UV excess quasar region’ $g-i$ values comprised between -0.2 and $+0.8$ and $\log(f_x/f_r) \geq -1.2$). In this range of parameters, the mean $\log(f_x/f_r)$ of the spectroscopically identified sample appears slightly shifted by ~ 0.3 dex to lower values (i.e. 0.75 mag brighter for a given X-ray flux), compared to that of the photometric sample. Since the mean r magnitude of the corresponding spectroscopic and photometric-only groups are of 18.86 and 20.73 respectively, as a result of the necessarily brighter optical flux limit of the spectroscopic sample, this indicates that the photometric sample is dominated by a slightly more remote population of AGN, hence fainter in X-rays and in optical than the spectroscopic sample, albeit with a somewhat larger mean f_x/f_r . It can also be seen in Fig. 7 that these UVX spectroscopically identified quasars are the most energetic with X-ray luminosities in excess of 10^{44} erg/s. In a general manner, Fig. 5 shows that the spectroscopically identified sample of point-like objects covers the range of parameters populated

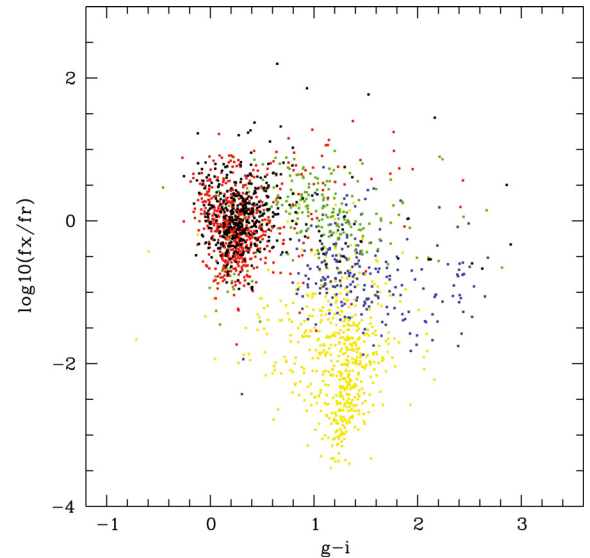


Fig. 7. Distribution of extragalactic objects in the $\log(f_x/f_r)$ versus $g-i$ diagram according to L_X . Black; $>10^{45}$, red 10^{44} to 10^{45} , green 10^{43} to 10^{44} , blue 10^{42} to 10^{43} , yellow 10^{38} to 10^{42} . We only show SDSS entries with a probability of identification with an X-ray source higher than 90%, g and i magnitudes brighter than 22.2 and errors on $g-i < 0.2$.

by the photometric cross-identifications for both AGN and stars relatively well, except, as quoted above, for the faintest optical matches. This identified sample could thus be used as a learning sample to statistically identify and classify X-ray sources with optical counterparts of comparable brightness.

This is at variance with the situation prevailing for extended sources. As seen in Fig. 6, a considerable number of X-ray sources are identified with red spatially extended photometric objects, i.e., relatively faint reddish galaxies with $g-i \gtrsim 1.0$ and corresponding $\log(f_x/f_r) \geq -0.5$. Unfortunately, the SDSS policy for selecting spectroscopic targets does not cover this region of the parameter space well. In the few cases in which an optical spectrum exists, they are assigned an AGN type. These

galaxies are significantly optically brighter than most UVX quasars, most of which are in the r mag range of 18 to 21. Their derived X-ray luminosities in the range of 10^{43-44} erg s⁻¹ (Fig. 7) clearly show that the vast majority of these reddish objects are likely Seyfert galaxies. This population extends downwards to lower f_x/f_r ratios, narrowing the $g-i$ range spanned and decreasing their X-ray luminosities. Eventually the brightest objects ($r \lesssim 18$) merge with the group of “normal” galaxies with $L_X \lesssim 10^{42}$ erg s⁻¹, which is well represented in the spectroscopically identified sample. These low X-ray luminosities could be explained in terms of ULXs, starbursts, or of a collection of low-mass X-ray binaries in elliptical galaxies.

8. Science cases

In the next two sections we touch upon two distinct science cases, one in the extragalactic domain and one related to a galactic source population. These two examples aim at illustrating the range of research that these clean cross-correlated samples allow and do not explore in depth all possible paths of investigations. In particular, we do not make use of the spectroscopic line data, which could provide many additional astrophysical diagnostics. The first case considered bears on the topical search for QSO2s, while the second one explores the X-ray and optical properties of active stellar coronae.

8.1. Searching for QSO2 candidates

The members of the high-luminosity high-obscuration part of the AGN population are commonly denominated QSO2s. The synthesis modelling of the XRB (Gilli et al. 2007; Treister et al. 2009) predict that up to ~20% to the XRB (Gilli et al. 2007) could be produced by QSO2s, they could represent ~30–50% of the high luminosity AGN population (e.g. Della Ceca et al. 2008), and they could probably co-evolve with massive host galaxies (Severgnini et al. 2006); it is therefore clear, that hunting for QSO2 remains one of the most topical activities as is the search for associated X-ray and optical signatures.

Optical selection of QSO2s relies on finding objects showing only narrow emission lines with high-ionisation line ratios and high luminosity typically from [OIII], e.g. $L_{\text{[OIII]}} > 10^{8.3} L_\odot$ (Zakamska et al. 2003; Reyes et al. 2008). X-ray selection looks instead for luminous ($L_{X,2-10\text{keV}} > 10^{44}$ erg/s) significantly obscured (column density $N_{\text{H}} > 10^{22}$ cm⁻²) sources, which are best selected in the $E > 2$ keV hard X-ray band (e.g. Mainieri et al. 2002; Caccianiga et al. 2004; Perola et al. 2004; Vignali et al. 2006; Della Ceca et al. 2008; Krumpe et al. 2008). Within the Unified Model (Antonucci 1993), obscuration of the central X-ray-emitting and Broad Line-emitting regions by an intervening torus gives rise to those consistent properties across both bands.

A priori, QSO2s should present red optical colours (since the emission of the host galaxy would dominate over the obscured AGN), high X-ray hardness ratios³ (because of the predominant absorption of the lower energy X-rays) and high X-ray-to-optical flux ratio (f_x/f_r , since the X-rays are less sensitive to absorption

than the optical range). However, several effects could alter this simple recipe. For instance, Compton Thick absorption (defined here as $N_{\text{H}} \gtrsim 10^{24}$ cm⁻²) would completely absorb direct X-rays up to 10 keV, this would alter both the spectral shape (since scattered primary X-rays have “softer” spectra) and the ratio of optical to X-ray fluxes would be more typical of “normal” galaxies (i.e. $\log(f_x/f_r) < -1$).

We investigate here whether the position in the overall X-ray and optical parameter space of the confirmed/candidates QSO2 discovered so far could give some hint on the way other candidates could be selected on the basis of broadband high-energy XMM-Newton data and optical photometry only.

To do so we first assembled a sample of *bona fide* QSO2 from optical and X-ray surveys. The optically selected sample has been obtained by cross-correlating the SDSS sample of Zakamska et al. (2003) and Reyes et al. (2008) with the 2XMMi catalogue. We list in Table 9 the main properties of the QSO2 SDSS candidates matching a serendipitous EPIC source. The listed X-ray luminosities were computed assuming an average shape (see Sect. 6.2) for the large band 0.2 to 12 keV energy distribution and are not corrected for intrinsic absorption⁴. We also marked in the table the eight SDSS QSO2 that are good candidates to be Compton Thick. To define these objects we used $L_{X,\text{meas}}/L_{X,\text{O[III]}} < 0.05^5$, where $L_{X,\text{meas}}$ is the observed 2–10 keV luminosity (see above) and $L_{X,\text{[OIII]}}$ is the expected intrinsic X-ray luminosity (the latter has been computed using the observed $L_{\text{O[III]}}$ and the ratio $L_{\text{O[III]}}/L_x \sim 0.017$ derived for the unobscured view of Seyfert galaxies, Heckman et al. 2005).

To this sample of optically selected QSO2 we added a small sample of five X-ray selected QSO2 (with the definition above) obtained by cross-correlating SDSS, 2XMMi, and a few selected lists of X-ray defined QSO2 (Della Ceca et al. 2008; Krumpe et al. 2008; Corral et al., in prep.). It is worth stressing that based on a detailed analysis of the X-ray and optical spectral properties (see Della Ceca et al. 2008; Corral et al., in prep.), all these X-ray defined QSO2 are Compton Thin (intrinsic N_{H} between 10^{22} cm⁻² and few times 10^{23} cm⁻²).

The position of the confirmed/candidates QSO2 in the parameter space obtained using f_x/f_r , the optical colours (in particular, $g-i$) and hardness ratio (in particular HR2) are shown in Figs. 8 and 9; we marked with different symbols the several “flavours” of QSO2, i.e. the X-ray selected QSO2 (all Compton Thin), the optically selected QSO2 and the candidate Compton Thick QSO2.

As can be seen in Fig. 9, the QSO2 generally appear slightly redder than the bulk of the unresolved AGN, and of the identified galaxy sample having $L_X > 10^{44}$ erg s⁻¹. Their colours are more similar to that of the galaxy identified sample with $10^{42} < L_X < 10^{44}$ erg s⁻¹ (green points in Fig. 9, left panel).

The position of the QSO2 in Figs. 8 and 9 clearly shows a separation between the “confirmed” Compton Thin QSO2 (magenta filled squares and triangles) and the “candidates” Compton Thick QSO2 (encircled). Interestingly, the four “optically selected” QSO2 occupying the same region of the X-ray selected Compton Thin QSO2 have been studied in the X-ray domain by (Ptak et al. 2006); all these sources

³ The XMM-Newton EPIC hardness ratios use five energy bands, which expressed in keV units are 0.2–0.5, 0.5–1, 1–2, 2–4.5 and 4.5–12.0. Hardness ratio i is expressed as

$$\text{HR}_i = \frac{C_{i+1} - C_i}{C_{i+1} + C_i}, \quad (17)$$

with C_i the count rate in band i corrected for vignetting.

⁴ These luminosities are necessarily less accurate than those derived by Ptak et al. (2006) from a detailed spectral analysis; however the four candidates in the list of Zakamska et al. (2003) in common with the eight discussed in Ptak et al. (2006) (all Compton Thin QSO2) have luminosities consistent within a factor of two.

⁵ Our criteria to define a Compton Thick AGN is slightly less restrictive than that recently used by e.g. Vignali et al. (2010), $L_{X,\text{meas}}/L_{X,\text{O[III]}} < \sim 0.01$.

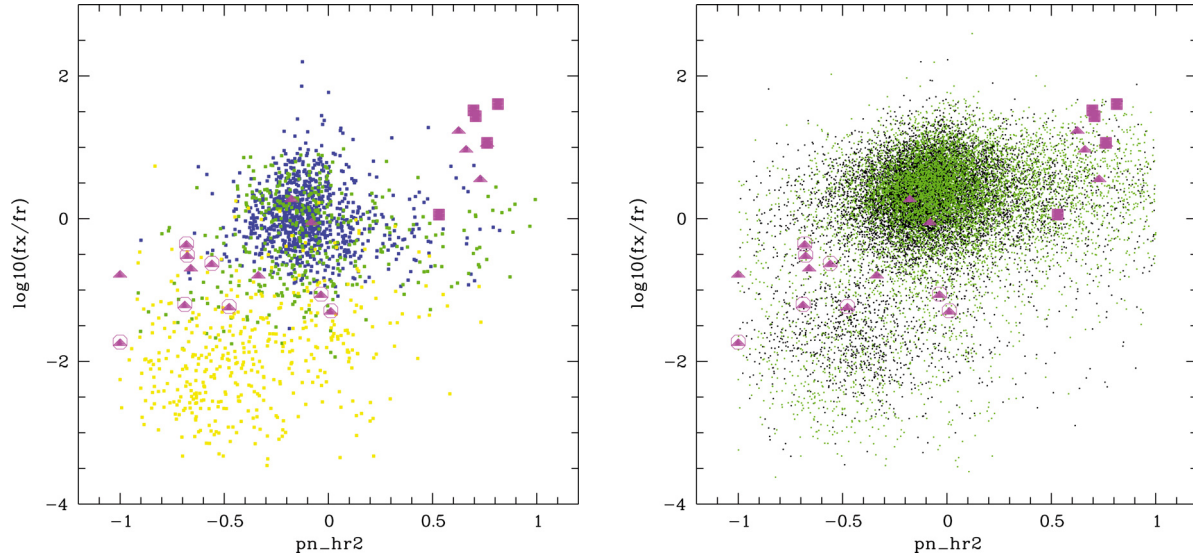


Fig. 8. *Left panel:* distribution of spectroscopically identified objects (optically resolved and unresolved) in the $\log(f_x/f_r)$ versus EPIC pn HR2 diagram. blue: AGNs with $\log(L_X) \geq 44$, green: $44 \geq \log(L_X) \geq 42$, yellow: galaxies with $\log(L_X) \leq 42$, magenta: type 2 QSO – filled squares = X-ray selected – filled triangles = optically selected – encircled = Compton Thick. *Right panel:* the entire SDSS photometric sample. Black: unresolved objects, green: extended objects, magenta: same as in right panel. We only show SDSS entries with a probability of identification with an X-ray source higher than 90%, r magnitudes brighter than 23 and errors on HR2 less than 0.3.

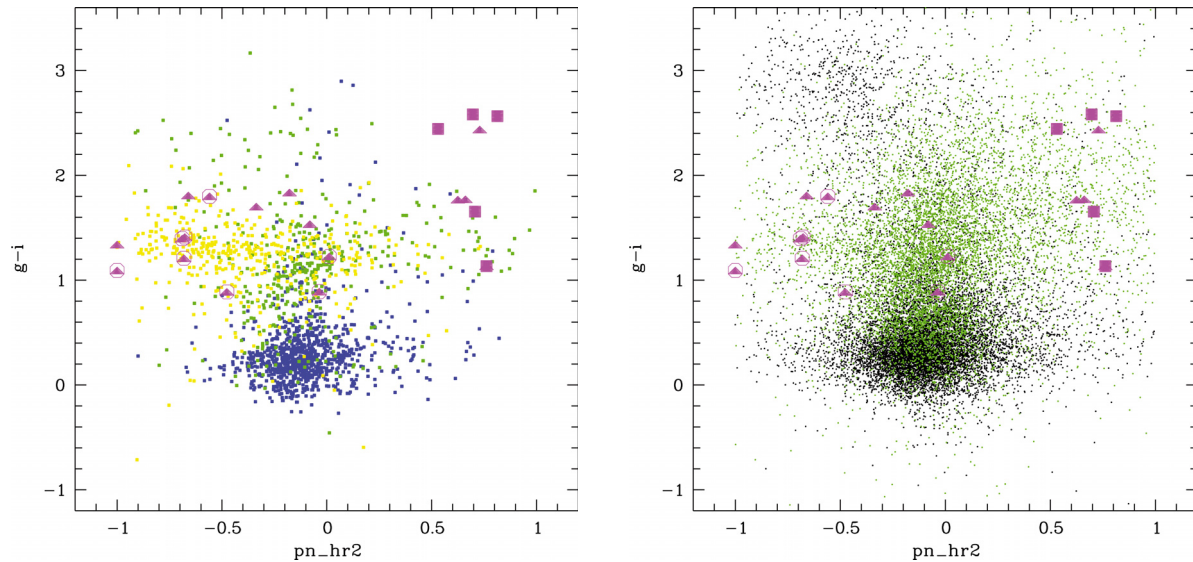


Fig. 9. *Left panel:* distribution of spectroscopically identified objects (optically resolved and unresolved) in the $g-i$ versus EPIC pn HR2 diagram. blue: AGNs with $\log(L_X) \geq 44$, green: $44 \geq \log(L_X) \geq 42$, yellow: galaxies with $\log(L_X) \leq 42$, magenta: Type 2 QSO – filled squares = X-ray selected – filled triangles = optically selected – encircled = Compton Thick. *Right panel:* the entire SDSS photometric sample. Black: unresolved objects, green: extended objects, magenta: same as in right panel. We only show SDSS entries with a probability of identification with an X-ray source larger than 90%, r magnitudes brighter than 23 and errors on HR2 less than 0.3.

(2XMMiJ011522.2+001518, 2XMMiJ021047.0–100152, 2XMMiJ122656.4+013124, 2XMMiJ164131.6+385841) are described by an absorbed power-law model with an intrinsic $N_H \sim 2-3 \times 10^{22} \text{ cm}^{-2}$. Therefore the upper right corner of Fig. 8 is probably the best place where to look for Compton Thin QSO2; as shown in (e.g. Caccianiga et al. 2004) the very positive HR2 reflects the relatively large intrinsic photoelectric absorption present in many QSO2 and responsible for their preferential discovery in hard X-ray surveys⁶.

⁶ We caution however that the separation in HR2 properties between Compton Thin and Compton Thick QSO2 is probably clear only for sources with redshift as high as in our learning sample ($z < \sim 0.8$); at

Finally, we show in Fig. 10 the behaviour of the EPIC pn HR2 hardness ratio with X-ray luminosity for all spectroscopic SDSS targets. The bulk of the SDSS type 1 QSOs with X-ray luminosities higher than $10^{44} \text{ erg s}^{-1}$ (0.2–12 keV) cluster around an hardness ratio $\langle \text{HR2} \rangle = -0.12$; the same objects cluster around hardness ratios $\langle \text{HR3} \rangle = -0.38$ and $\langle \text{HR4} \rangle = -0.28$. These hardness ratios are in excellent agreement with the values expected from a canonical $\Gamma = 1.9$ power law X-ray spectrum undergoing negligible intrinsic absorption and a mean Galactic absorption of $1.16 \times 10^{20} \text{ cm}^{-2}$ (the average over all directions

higher redshift we should measure decreasing HR2 values for Compton Thin QSO2 because the observed X-ray energy band will move to higher energies, which are increasingly less affected by absorption.

Table 8. Main properties of the 2XMMi/SDSS identifications with $\log(f_x/f_r) \geq -0.5$ in each 0.5 $g-i$ bin.

$g-i$	Ntot	r mag (rms)	pn HR2 (rms)	pn HR3 (rms)	pn HR4 (rms)
point-like sources					
+0.25	9538	+20.20 (1.07)	-0.115 (0.228)	-0.397 (0.237)	-0.219 (0.376)
+0.75	3034	+20.59 (0.99)	-0.045 (0.250)	-0.360 (0.246)	-0.199 (0.330)
+1.25	471	+20.51 (0.84)	+0.061 (0.352)	-0.294 (0.317)	-0.188 (0.338)
+1.75	96	+20.17 (0.97)	+0.168 (0.280)	-0.229 (0.294)	-0.150 (0.347)
+2.25	31	+19.79 (1.02)	+0.200 (0.315)	-0.307 (0.282)	-0.168 (0.598)
+2.75	29	+19.04 (1.31)	-0.081 (0.368)	-0.447 (0.302)	-0.162 (0.545)
extended sources					
+0.25	812	+20.97 (0.84)	-0.099 (0.222)	-0.385 (0.255)	-0.175 (0.425)
+0.75	1915	+20.23 (1.46)	-0.100 (0.237)	-0.370 (0.235)	-0.246 (0.311)
+1.25	1918	+19.77 (1.44)	-0.009 (0.307)	-0.285 (0.335)	-0.149 (0.341)
+1.75	1364	+19.89 (0.97)	+0.172 (0.377)	-0.111 (0.452)	-0.028 (0.383)
+2.25	479	+20.00 (0.60)	+0.233 (0.414)	-0.012 (0.482)	-0.018 (0.337)
+2.75	35	+19.81 (0.78)	+0.027 (0.456)	+0.056 (0.499)	+0.027 (0.395)

Notes. We only consider here sources with error on $g-i \leq 0.2$, g and i mags brighter than 22.2 and individual probability of identification greater than 90%.

Table 9. QSO2 candidates from Zakamska et al. (2003), Reyes et al. (2008) and X-ray selected samples matching 2XMMi entries with EPIC pn observations.

2XMMi name	id prob	z	$u-g$	$g-i$	HR 2 (pn)	$\log(L_X)$	$\log(f_x/f_r)$
Zakamska et al. (2003)							
2XMM J005621.6+003235 ^{Tck}	0.964	0.484	1.891 ± 0.643	1.800 ± 0.061	-0.56 ± 0.28	42.78	-0.626
2XMMi J011522.2+00151	0.997	0.390	6.023 ± 1.239	2.437 ± 0.046	0.73 ± 0.04	44.26	0.566
2XMM J015716.9-005305	0.988	0.540	0.565 ± 0.167	1.836 ± 0.050	-0.18 ± 0.46	42.78	0.285
2XMM J021047.0-100152	0.999	0.540	0.348 ± 0.136	1.771 ± 0.047	0.66 ± 0.12	44.40	0.980
2XMM J103951.5+643005 ^{Tck}	0.991	0.402	0.490 ± 0.075	1.209 ± 0.027	-0.68 ± 0.33	43.02	-0.353
2XMM J122656.4+013124	0.998	0.732	0.970 ± 0.256	1.134 ± 0.056	0.76 ± 0.06	44.83	1.063
2XMM J164131.6+385841	0.995	0.596	0.550 ± 0.112	1.768 ± 0.028	0.63 ± 0.05	44.96	1.245
Additional candidates from Reyes et al. (2008)							
2XMMi J075821.2+392337 ^{Tck}	0.912	0.216	0.620 ± 0.026	0.891 ± 0.010	-0.04 ± 0.27	42.37	-1.058
2XMM J093952.7+355358	1.000	0.137	1.527 ± 0.051	1.534 ± 0.007	-0.08 ± 0.06	43.44	-0.040
2XMM J094506.4+035552 ^{Tck}	0.995	0.156	0.979 ± 0.043	1.092 ± 0.025	-1.00 ± 0.28	41.63	-1.730
2XMM J100327.8+554155	0.991	0.146	1.390 ± 0.066	1.340 ± 0.010	-1.00 ± 0.17	42.40	-0.772
2XMM J103408.5+600152 ^{Tck}	1.000	0.051	1.408 ± 0.009	0.888 ± 0.003	-0.48 ± 0.05	42.29	-1.232
2XMM J103456.3+393939 ^{Tck}	0.995	0.151	1.465 ± 0.049	1.412 ± 0.007	-0.68 ± 0.29	42.88	-0.512
2XMM J122709.8+124855	0.998	0.194	1.728 ± 0.162	1.808 ± 0.010	-0.66 ± 0.27	42.88	-0.686
2XMM J131104.6+272806	0.998	0.240	1.272 ± 0.070	1.702 ± 0.010	-0.34 ± 0.07	42.94	-0.784
2XMM J132419.8+053704 ^{Tck}	0.997	0.203	0.441 ± 0.027	1.226 ± 0.009	0.01 ± 0.47	42.24	-1.290
2XMM J171350.7+572955 ^{Tck}	0.883	0.113	1.279 ± 0.025	1.396 ± 0.006	-0.69 ± 0.21	42.10	-1.201
X-ray selected QSO2							
2XMM J113148.6+311400	0.995	0.50	0.724 ± 0.491	1.654 ± 0.091	0.71 ± 0.16	44.70	1.437
2XMM J122656.4+013124	0.998	0.73	0.970 ± 0.256	1.134 ± 0.056	0.76 ± 0.06	44.83	1.063
2XMM J134656.6+580316	0.965	0.37	1.472 ± 0.538	2.443 ± 0.069	0.53 ± 0.23	43.88	0.057
2XMM J160645.9+081523	0.960	0.62	0.919 ± 1.085	2.565 ± 0.216	0.81 ± 0.26	44.80	1.607
2XMM J204043.2-004548	0.336	0.62	1.448 ± 2.334	2.581 ± 0.304	0.70 ± 0.11	44.72	1.518

Notes. The source 2XMM J122656.4+013124 is present both in the optically selected and in the X-ray selected QSO2 sample.

of galaxy and QSO targets). For this group of QSOs, there is no evidence of strong dependence of the power-law index with L_X . There is however a small number of QSOs exhibiting a considerably harder X-ray spectrum (as testified by an increasing value of HR2) extending to the same locus occupied by confirmed Compton Thin QSO2. As one enters the AGN regime at X-ray luminosities below 10^{44} erg s⁻¹, the number of extended sources with galaxy-like optical spectra rises considerably and the shape of the X-ray energy distribution shows a much larger scatter. The candidates Compton Thick QSO2 seem to populate this part of the diagram.

A number of spectroscopic SDSS entries occupy the same region of the LX/HR2 diagram as the reference Compton Thin

QSO2. We explored the nature of these candidates by selecting objects with $\log(L_X)$ higher than 44, EPIC pn HR2 greater than 0.5 with an error of less than 0.2 on the hardness ratio: twelve objects match these conditions. From an inspection of the spectroscopic SDSS data and a literature search we found that at least ~60% of them are indeed characterised by absorption at same level: two objects are clearly Broad Absorption Line QSOs (SDSS J114312.32+200346.0, SDSS J141546.24+112943.4), two are “dust reddened QSOs” (SDSS J122637.02+013016.0, SDSS J143513.90+484149.2) and three are Type 2 QSOs (SDSS J105144.24+353930.7, SDSS J130005.34+163214.8, SDSS J134507.93-001900.9). The remaining five objects are apparently “normal” type I AGN,

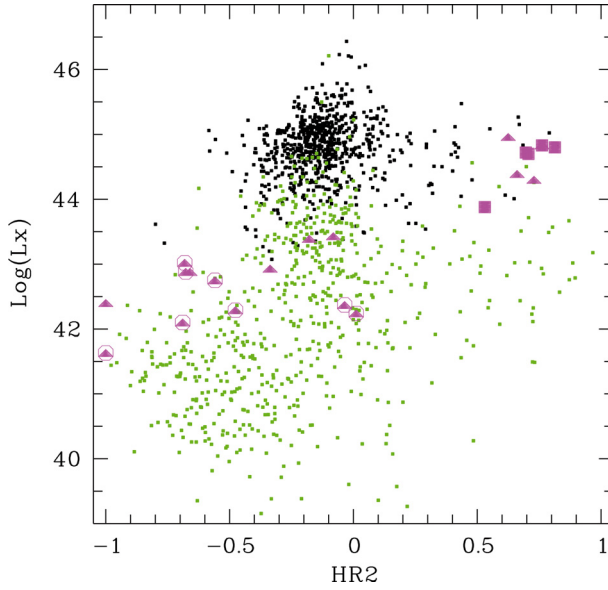


Fig. 10. Variation of the EPIC pn hardness ratios HR2 with X-ray luminosity for extragalactic spectroscopic SDSS targets optically unresolved (black dots) and optically extended (green dots). Magenta: type 2 QSO – filled squares = X-ray selected – filled triangles = optically selected – encircled = Compton Thick. Errors on hardness ratios are below 0.2 in all cases.

without any specific comment in literature: a detailed analysis of their optical and X-ray properties (e.g. to understand if these latter objects could also be classified as “dust reddened QSOs”) is beyond the scope of the present paper. The main results of this exploration is that, although Compton Thin QSO2 do separate well in the LX/HR2 diagram, other rare kinds of objects occupy the same locus.

Finally we investigated the possible presence of Compton Thin QSO2 candidates in the SDSS photometric matches. We did this by looking for sources showing evidences of photoelectric absorption among the 2XMMi sources matching mainly optically extended SDSS objects characterised by a high f_x/f_r ratio and relatively red $g-i$ colours. We thus built histograms of the EPIC pn hardness ratios for all 2XMMi/SDSS matches with $\log(f_x/f_r) \geq -0.5$ for various $g-i$ colour intervals. The resulting histogram for the EPIC pn HR2 hardness ratio is shown in Fig. 11.

The histograms clearly show a main peak of hardness ratio corresponding to the canonical low $N_H \Gamma = 1.9$ power-law spectrum characteristic of the type I QSOs for all $g-i$ colour indexes. However, for the reddest objects, typically with $g-i \geq 1.5$, a secondary bump is observed for harder hardness ratios ($HR2 \approx 0.5$) with values consistent with those of the Compton Thin QSO2; a similar secondary bump is also present if we use the hardness ratios HR3 ($HR3 \approx 0.35$) and HR4 ($HR4 \approx 0.15$). A detailed investigation of these possibly interesting sources is beyond the scope of the present paper.

8.2. X-ray active stars

At the high galactic latitudes covered here by the legacy SDSS survey ($|b| \gtrsim 20^\circ$ with a mean $|b|$ of 58°), most of the X-ray sources are of extragalactic origin. For instance, the Extended Medium Sensitivity Survey, which constitutes the largest optically identified sample of serendipitous Einstein X-ray sources at high galactic latitude, contains only 25% of active stars

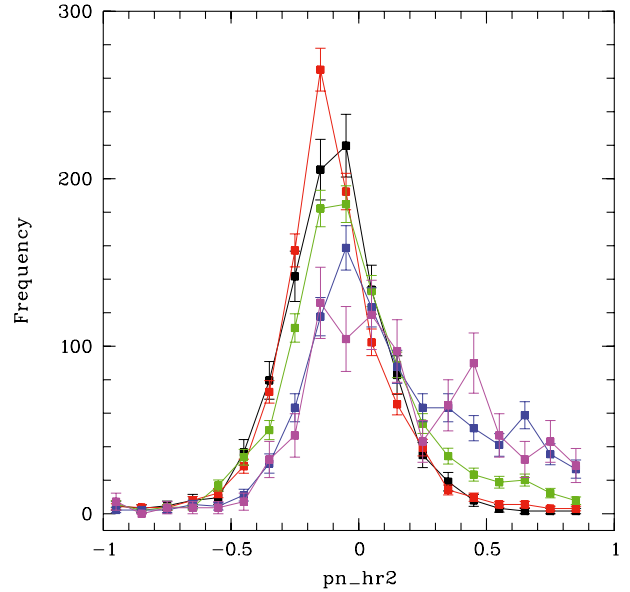


Fig. 11. Variation of the EPIC pn hardness ratio HR2 with $g-i$ colour for all 2XMMi matches with extended sources in the entire SDSS catalogue only having a $\log(f_x/f_r)$ larger than -0.5 . Each histogram corresponds to a $g-i$ colour bin of 0.5. Black; $g-i = 0.25$, red; 0.75, green; 1.25, blue; 1.75, magenta; 2.25. Errors on hardness ratios are below 0.3 in all cases. We only show identifications with an individual identification probability higher than 90%, g and i mag brighter than 22.2, errors on $g-i$ less than 0.2. Histograms are normalised to the total number of sources in the colour bins shown in Table 8.

(Stocke et al. 1991). Optical identification campaigns of the ROSAT all-sky survey sources (RASS, see Schwobe et al. 2000, for the bright sample) yielded similar results (e.g. 35% of active stars in Zickgraf et al. 1997), while the ROSAT deep survey of the Lockman field (Schmidt et al. 1998), which is ~ 50 times more sensitive than the RASS, collected less than 10% of stellar sources. Stars constitute a bounded population of comparatively soft X-ray sources. Consequently, their relative contributions to the high galactic source number count is expected to decrease very significantly with increasing sensitivity in Chandra and XMM-Newton observations, which both offer a lower flux limit and harder energy response (see for instance Fig. 1 in Hérent et al. 2006). Active coronae indeed constitute less than 10% of all sources identified in the XMM-Newton serendipitous survey of Barcons et al. (2007). The deepest Chandra surveys such as the Chandra Deep Field North (Alexander et al. 2003) counts only $\sim 3\%$ of stars.

This is at variance with the low latitude situation where most of the RASS X-ray sources were found to be associated with stars (Motch et al. 1997). XMM-Newton and Chandra have extended to lower fluxes the predominance of active coronae on the soft X-ray low latitude source population (Motch et al. 2003; Rogel et al. 2006). At higher energies, a longitude dependent population of galactic hard X-ray sources appears on the top of a usually dominant background of extragalactic sources (Hands et al. 2004; Motch 2006; Motch et al. 2010).

As stated above, the number of X-ray emitting and spectroscopically identified stars available in the DR7 is relatively small, and because of the scientific goals put forward at the time of the selection of targets for spectroscopic follow-up, concentrates on the reddest M type stars. The availability of the SEGUE archive in DR7 has somewhat increased the number of spectroscopically identified stars, but its effect on the cross-correlation

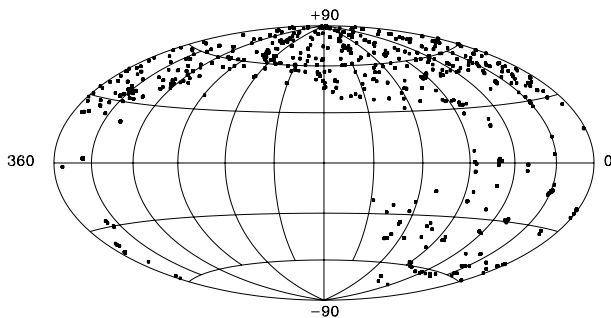


Fig. 12. Aitoff projection of the distribution of the 2XMMi sources matching identified stars with a probability $\geq 90\%$.

statistics remains small. As mentioned in Sect. 6, in order to increase the stellar sample towards earlier types, we used a kernel density classification to identify the SDSS DR7/2XMMi matches with multicolour properties consistent with those expected from stars of main sequence class.

Figure 12 shows the distribution in galactic coordinates of all X-ray active stars present in the identified sample. Including SEGUE data in the DR7 has allowed the identification of a few stars at low galactic latitude. However, the mean $|b|$ of the stellar sample remains high, ($|b| \sim 45^\circ$) and therefore does not change the conclusion that the present stellar sample is typical of the high Galactic latitudes.

At high b and typical distances of a few hundred parsecs, interstellar absorption remains negligible compared to other uncertainties and has little effect on the observed stellar colours. We computed the total galactic reddening in the directions of each of the X-ray emitting stars following Schlegel et al. (1998). The average $E(B-V)$ is 0.029 with a rms of 0.025. Assuming the absorption coefficients computed by (Girardi et al. 2004) ($T_{\text{eff}} = 4500$ K; $\log g = 4.5$), the maximum reddening applicable on average to our sample of stars would be of 0.028 and 0.018 in the $g-r$ and $r-i$ colours respectively. Only very few spatially unresolved SDSS sources matching 2XMMi entries have combined $u-g$, $g-r$ and $r-i$ colours compatible with those expected from giant class III stars, not to mention supergiants. In a recent paper, Guillout et al. (2009) estimate that their sample of $|b| \lesssim 30^\circ$ RASS sources identified with bright Tycho stars has a mean contamination of 35% by evolved stars with a peak at 60% for K stars. However, a smaller fraction of X-ray emitting evolved stars of $\sim 10\%$ was present in the sample of Covey et al. (2008), which is more representative of what we should expect in our case because it was selected at higher galactic latitude and fainter flux. Below, we will assume that all stars belong to the main sequence, keeping in mind that a fraction of the class III and class IV stars, in particular in short period binaries such as RS CVn systems, could contribute to some extent. Being considered as single dwarfs, these stars would have computed X-ray luminosities below that actually emitted. Our sample comprises 549 active coronae candidates with individual probability of identification above 90%, corresponding to a total sample reliability of 98%. The interval of $g-r$ and $g-i$ colours corresponds to K4 to late M5-M6 stars. Earlier stars are more often subject to optical saturation and have in general lower KDC probabilities. They are therefore excluded from the clean identified stellar sample built here.

Neglecting reddening effects, we computed the distances using the absolute magnitude calibration listed in Covey et al. (2007) and X-ray luminosity using the broad band (0.2–12 keV) flux listed in the 2XMMi catalogue for the EPIC camera. The

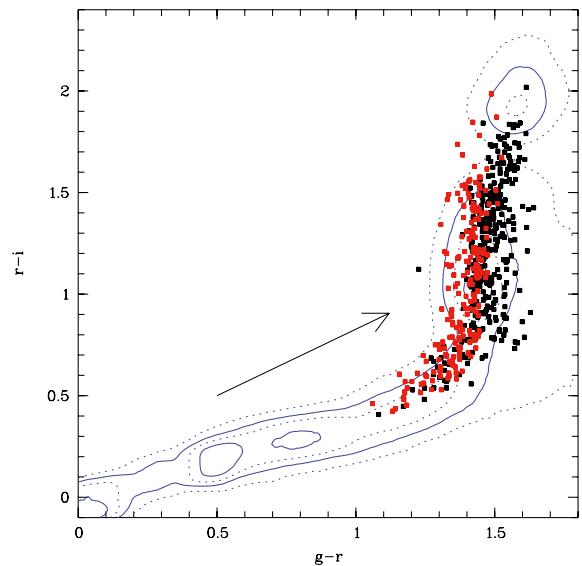


Fig. 13. Distribution of X-ray active stars in the $g-r/r-i$ diagram for two ranges of X-ray luminosities. Black dots $\text{Log}(L_x) < 29.3$. Red crosses $\text{Log}(L_x) > 29.3$. The arrow shows the direction of interstellar reddening. Only sources with photometric errors below 0.1 mag in $g-r$ and $r-i$ are shown. Density contours show the position of all stars identified in the SEGUE programme.

mean photometric distance is 340 pc and all stars have distances in the range of 40 pc to 2000 pc. The overall $\text{Log}(L_x)$ (0.2–12 keV) distribution peaks at 29.20 and ranges from ~ 27.1 to ~ 30.9 . This interval of X-ray luminosity covers that exhibited by old stars such as the Sun or even less active, up to that emitted by the most active T Tauri or RS CVn stars. The mean X-ray luminosity does not vary with galactic latitude in the interval covered by our sample.

We show in Fig. 13 the distribution of X-ray active stars in the $g-r/r-i$ diagram for two ranges of X-ray luminosities. It can be readily seen that the locus of X-ray bright sources is shifted by ~ 0.1 mag in colour above that occupied by low L_x active coronae and by spectroscopically identified stars from SEGUE in general. Active stars appear to be bluer in $g-r$ for a given $r-i$. The effect is particularly clear for the reddest stars with $g-r \gtrsim 1.3$, corresponding to M0 types and later. This shift cannot be due to the lack of reddening correction since the most X-ray luminous stars that are expected to be the most remote and absorbed ones should appear redder than the closest low L_x stars, a trend opposite to what is observed. We also checked with the isochrones of Girardi et al. (2004) designed for the SDSS band passes that age or metallicity effects were unable to explain the different colour/colour tracks of the low and high L_x stars. Similarly, an enhanced $H\alpha$ emission expected to be especially important for late M stars ($g-r > 1.3$) can be excluded since it would yield larger $g-r$. Interestingly, Covey et al. (2008) report that the counterparts of their extended ChaMP stellar X-ray survey do exhibit a ~ 0.1 bluer $u-g$ index than average low-mass stars, although they do not report any significant change in the $g-r$ colour index. Comparing with their work is however not straightforward. While the colour/colour track followed on average by our active coronae is consistent with that of Covey et al. (2007) for the entire SDSS and therefore does not indeed contradict the results of Covey et al. (2008), the difference we find rather arises among two groups of active stars. The SDSS M stars templates compiled by Bochanski et al. (2007) could indicate a similar trend in the $u-g$ colour index between active

and inactive stars with however a large intrinsic scatter, while no such effect occurs in $g-r$. Unfortunately, we have too few good u band measurements to be able to confirm the $u-g$ trend seen by Covey et al. (2008). We agree with these authors that low-level optical flaring might be responsible for the bluer colours seen in X-ray active M dwarfs.

Active coronae emit essentially thin thermal spectra dominated by a series of narrow emission lines superposed on a weak continuum. In most cases, two thermal components are required to satisfactorily represent the observed energy distribution (see a recent review in Güdel & Nazé 2009). X-ray studies of open clusters and field stars of different ages led to a relatively coherent picture linking stellar rotation rates, overall X-ray luminosity and X-ray temperatures. Whereas the young (age < 1 Myr) stars in Orion exhibit X-ray spectra with $kT_1 \sim 0.8$ keV and $kT_2 \sim 2.9$ keV, the analysis of ~ 115 Myr old Pleiades stars yields $kT_1 \sim 0.4$ keV and $kT_2 \sim 1.1$ keV, while the X-ray corona of our Sun can also be characterised by 2T spectrum with $kT_1 \sim 0.2$ keV and $kT_2 \sim 0.6$ keV (see Sung et al. 2008, and references therein). Using ROSAT and ASCA observations of a dozen of carefully selected stars, Guedel et al. (1997) established that the overall X-ray luminosity, the temperature of the two components, and the emission measurement ratio of the hot to the cool plasma were all decreasing with age and rotation rate. The range of X-ray luminosity observed in our survey suggests ages younger than ~ 2 Gyr.

We thus investigated whether the X-ray properties of our identified active coronae were depending on luminosity. For that purpose we selected the 296 (149) identified stellar X-ray sources having errors on 2XMMi hardness ratio 2 (3) of less than 0.2. This sample was then split into five ranges of luminosity for which the mean HR2 was computed, and in order to accommodate the lower number of sources in the harder band, into three bands of X-ray luminosities for HR3. Hardness ratio 2 and 3 measure the relative count rates in the energy ranges 0.5–1.0 keV and 1.0–2.0 keV and 1.0–2.0 keV and 2.0–4.5 keV respectively. They are therefore weakly dependent on the EPIC filter used and vary little within the range of N_H applicable to the present survey. HR2 and HR3 can thus be considered as indicators of the intrinsic shape of the X-ray energy distribution well suited to the range of temperatures exhibited by stellar coronae. We show in Fig. 14 the variation of the median HR2 and HR3 with X-ray luminosity. A clear spectral hardening accompanies the luminosity increase. Assuming a single temperature plasma undergoing $N_H \sim 1.7 \times 10^{20} \text{ cm}^{-2}$ the HR2 value would imply a thin thermal temperature of ~ 0.45 for the lowest L_X bin and 0.72 keV for the largest L_X , while for the same range of X-ray luminosities the HR3 value would indicate temperatures from ~ 0.45 keV to ≥ 1.4 keV. The existence of a second hotter thin thermal component naturally accounts for this discrepancy. Unfortunately, using only two hardness ratios, it is impossible to fit both kT_1 , kT_2 and the ratio of the emission measurements of the hot and cool components, which is also expected to vary with X-ray luminosity (Guedel et al. 1997). In spite of this shortcoming, our data which benefit from the large throughput of XMM-Newton and of its capability to obtain detailed spectral information for an unprecedented number of sources, confirm for the first time the spectral hardening with X-ray luminosity in large groups of mixed age field objects.

9. Summary

We present the results of the cross-identification of the 2XMMi catalogue containing over 220 000 XMM-Newton EPIC

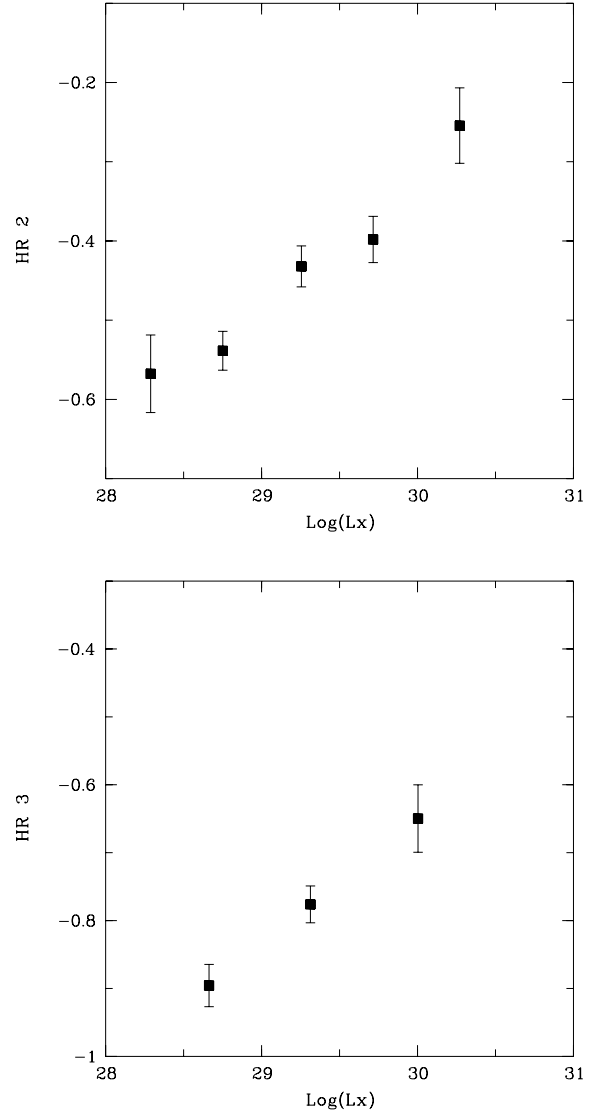


Fig. 14. Variation of the EPIC pn Hardness ratios 2 and 3 with mean X-ray luminosity for stellar X-ray sources with HR errors below 0.2. We plot here the median and associated errors.

serendipitous sources with the Data Release 7 of the Sloan Digital Sky Survey consisting of 357 million unique objects and over 1.6 million spectra. In order to ensure the best statistical control on the quality of the cross-correlation process, we use a likelihood ratio scheme only based on the probability of spatial coincidence of the X-ray source with the optical candidate. Using an original method that does not rely on heavy Monte Carlo simulations, we are able to compute true probabilities of identifications taking into account the varying fraction of X-ray sources expected to have a SDSS match as function of galactic latitude. We discuss the details of the statistical method used and present reliability and completeness curves for the entire set of matches. A total of 30 000 unique X-ray sources have a match in the DR7 with a probability of identification above 90%. At this threshold, the reliability of the total sample is of 98%, i.e. we expect only 2% of spurious cross-identifications, while the completeness is 77%, meaning that we miss about a quarter of the true matches, which therefore appear with individual probabilities of identifications below the 90% threshold.

One of the most ambitious task attributed to the Survey Science Centre of the XMM-Newton satellite is the statistical

identification of all serendipitous X-ray sources discovered in the large field of view of the EPIC cameras. The 2XMMi/DR7 cross-identified sample offers an interesting opportunity to test various strategies and methods which would eventually led to the classification of the XMM catalogues. Pineau et al. (2008) presented a first attempt to reduce the parameter space dimension using Principal Component Analysis tools adapted to data affected by rather large instrumental errors and taking into account the intrinsic scatter of the measured quantities. These authors also tested the relative merits of various classification methods. We use the SDSS DR7 spectroscopic catalogue to build an identified sample made of objects of known nature to which we add a few entries extracted from specific additional catalogues (e.g. Cataclysmic Variables). We find that the most discriminating diagrams involve the f_x/f_r flux ratio and various other parameters such as the $g - i$ colour index, EPIC hardness ratios, and the extent of the optical source. Active galactic nuclei more luminous than 10^{42} erg s⁻¹ have f_x/f_r ratios well above those of most galactic stars in their range of $g - i$. Galaxies of lower X-ray luminosity overlap the region of the diagram covered by active coronae, but can be easily separated from them on the basis of the optical extent.

Not unexpectedly, active galactic nuclei constitute by far the most numerous class of X-ray emitters present in the cross-correlation. Optically unresolved UV excess quasars are the most frequent X-ray emitters encountered and cluster in a well defined area in the $(f_x/f_r)/(g - i)$ diagram. The second largest group of X-ray detected AGN consists of optically extended objects, with similar or only slightly lower f_x/f_r ratios, but exhibiting considerably redder $g - i$ colours. The very scarce SDSS spectroscopic coverage of this region does not allow constraining their true nature. However, relatively bright galaxies harbouring active nuclei with L_X somewhat lower than those of UV excess quasars are likely to account for most of these objects. The vast majority of these “red” AGN exhibit X-ray hardness ratios consistent with those expected from a typical type I nucleus. However, a small fraction of these sources display evidence of enhanced photo-electric absorption that could be the signature of Compton Thin QSO2s, although other species of AGN can overlap the same parameter region. As L_X (and f_x/f_r) decrease, extragalactic X-ray sources become spatially resolved in the optical and their $g - i$ colour indices tend to cluster around $g - i \sim 1.3$. In parallel, the hardness ratios of lower L_X sources appear globally softer. We probably see here the effect of the increasing contribution of non-nuclear X-ray sources. Although some stars share f_x/f_r ratios and $g - i$ colours similar to some galaxies, taking into account the extent of the optical object allows an almost perfect separation of the galactic and extragalactic source populations. As expected, the latest M stars display larger f_x/f_r ratios than earlier active coronae.

Using a sample of reliable X-ray and optical Compton Thin and Compton Thick QSO2s we investigated the possibility of preferentially finding these objects in some specific range of X-ray and optical parameters. We find that Compton Thick QSO2 lack outstanding observational optical and X-ray photometric signatures that may be used to select sub-samples with greatly enhanced densities. However, the photo-electric absorption detectable in X-rays for relatively low-redshift AGN could be used to set apart groups of Compton Thin candidates with a reasonable success rate.

We also find that X-ray bright stars appear slightly bluer in $g - r$ than less active coronae. The origin of this effect is not clear, but could be related to the UV and blue micro flaring occurring in many of the active late type K and M stars. Assuming

that most active stars are indeed main sequence, we find a strong dependency of EPIC hardness ratios 2 and 3 with the inferred X-ray luminosity, the most luminous coronae being those emitting with the highest temperature.

The cross-correlation of two large catalogues such as the 2XMMi and the DR7 of the SDSS paves the way to a wide range of investigations. In addition, the SDSS database contains line fluxes for over a million of spectroscopic targets, while the 2XMMi catalogue provides X-ray spectra and time series for the 44 000 brightest serendipitous XMM-Newton sources. These rich data sets offer unprecedented opportunities for studying how the micro-physics of the optically emitting regions relates to X-ray properties. Eventually, such studies, in particular those aiming at finding rare outliers, rely on a solid control of the cross-correlation and cross-identification statistics.

We presented the details of an algorithm based on the classical likelihood ratio, well adapted to cross-correlation of large catalogues and providing useful probabilities of identification. This algorithm is now available as a plug-in for the Aladin VO portal⁷.

Acknowledgements. We thank an anonymous referee for providing constructive comments and help in improving the contents of this paper. F.-X. Pineau acknowledges support from CNRS, CNES and from the *Region Alsace*. R.D.C. acknowledges financial support from the Italian Space Agency (ASI) and from the Ministero Istruzione, Università e Ricerca (MIUR) during the past years. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington. This research has made use of the SIMBAD database and of the VizieR catalogue access tool, operated at CDS, Strasbourg, France.

Appendix A: Transformation of error ellipses in the new frame

A.1. Spherical trigonometry

On the unit sphere of centre C , we consider the spherical triangle NXO where N is the north pole, X and O are the positions of the X-ray and the optical sources respectively. We write

- x, n, o the angular distances between N and O , O and X , N and X respectively.
- \hat{N}, \hat{X} and \hat{O} the dihedral angles between planes NCX and NCO , XCN and XCO , OCN and OCX respectively. Those angles in the range $]-180^\circ, 180^\circ]$ and are define as positive in the East direction.

We immediately have $\hat{N} = \alpha_O - \alpha_X$, $x = 90^\circ - \delta_O$, $o = 90^\circ - \delta_X$ and $n = d$ the angular distance between the two sources, which

⁷ The plugin and its documentation can be downloaded from the home page of the XCat-DB (<http://xcatdb.u-strasbg.fr/>).

is given by the Haversine formula

$$d = 2 \arcsin \sqrt{\sin^2 \frac{\delta_o - \delta_X}{2} + \sin^2 \frac{\alpha_o - \alpha_X}{2} \cos \delta_X \cos \delta_o}. \quad (\text{A.1})$$

The spherical trigonometry gives the formula

$$\frac{\sin \hat{N}}{n} = \frac{\sin \hat{X}}{x} = \frac{\sin \hat{O}}{o}. \quad (\text{A.2})$$

As n , x and o are positive, we notice that \hat{N} , \hat{X} and \hat{O} have the same sign, which depends on $\alpha_o - \alpha_X$. As the function \arcsin returns values $\in [-90, +90]$, we deduce two possible values for both angle \hat{X} and \hat{O} . The cosine of half-angles $\hat{X}/2$ and $\hat{O}/2$ enable to know if $|\hat{X}| > 90^\circ$ and if $|\hat{O}| > 90^\circ$ respectively. From the spherical trigonometry formulae

$$s = \frac{1}{2}(n + x + o) \quad (\text{A.3})$$

$$\cos \frac{\hat{X}}{2} = \sqrt{\frac{\sin s \sin (s - x)}{\sin n \sin o}} \quad (\text{A.4})$$

$$\cos \frac{\hat{O}}{2} = \sqrt{\frac{\sin s \sin (s - o)}{\sin n \sin x}} \quad (\text{A.5})$$

We derive the values of angles \hat{X} and \hat{O} :

$$\hat{X} = \begin{cases} 180 - \arcsin \frac{x \sin \hat{N}}{n} & \text{if} \\ \cos \frac{\hat{X}}{2} > \frac{\sqrt{2}}{2} & \text{and } \arcsin \frac{x \sin \hat{N}}{n} \geq 0 \\ -180 - \arcsin \frac{x \sin \hat{N}}{n} & \text{if} \\ \cos \frac{\hat{X}}{2} > \frac{\sqrt{2}}{2} & \text{and } \arcsin \frac{x \sin \hat{N}}{n} < 0 \\ \arcsin \frac{x \sin \hat{N}}{n} & \text{else} \end{cases} \quad (\text{A.6})$$

$$\hat{O} = \begin{cases} 180 - \arcsin \frac{o \sin \hat{N}}{n} & \text{if} \\ \cos \frac{\hat{O}}{2} > \frac{\sqrt{2}}{2} & \text{and } \arcsin \frac{o \sin \hat{N}}{n} \geq 0 \\ -180 - \arcsin \frac{o \sin \hat{N}}{n} & \text{if} \\ \cos \frac{\hat{O}}{2} > \frac{\sqrt{2}}{2} & \text{and } \arcsin \frac{o \sin \hat{N}}{n} < 0 \\ \arcsin \frac{o \sin \hat{N}}{n} & \text{else.} \end{cases} \quad (\text{A.7})$$

A.2. Transformation of an error ellipse into a variance-covariance matrix

Our starting frame F_s is the plane centred in X and perpendicular in X to (CX) . It has as its x -axis the East direction and as its y -axis the North direction. Errors on positions in astronomy are almost always given in the form of an ellipse of semimajor axis a , semiminor axis b and of angle θ between the north pole and the major axis part oriented in the East direction. So $\theta \in [0, 180[$. We can transform the ellipse into a variance-covariance matrix in the frame F_s . The frame in which the ellipse is in its canonical form is obtain by rotating the frame F_s of an angle $\beta = 90 - \theta$. Applying a rotation of angle $-\beta$ on the canonical variance-covariance matrix give us the expression of the variance-covariance matrix in F_s :

$$V_{F_s} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix}. \quad (\text{A.8})$$

It leads to

$$\sigma_x^2 = a \sin^2 \theta + b \cos^2 \theta \quad (\text{A.9})$$

$$\sigma_y^2 = a \cos^2 \theta + b \sin^2 \theta \quad (\text{A.10})$$

$$\rho \sigma_x \sigma_y = \cos \theta \sin \theta (a - b). \quad (\text{A.11})$$

A.3. Variance-covariance matrices in our new frame

As mention in Sect. 3.1, our new frame F_n is centred in X , has as its x -axis the direction of the perpendicular in X of (CO) in the plane XCO oriented in the East part and as its y -axis the perpendicular in X of the plane XCO oriented in the North part. Given the results of the previous section, the new frame is obtained from the starting one (F_s) by a rotation of angle $\Omega = 90 - \hat{X}$ if \hat{X} is positive and $\Omega = -90 - \hat{X}$ if \hat{X} is negative. The variance-covariance matrix of the errors on positions can be expressed in the new frame F_n by a matrix V_n

$$V_n = \begin{pmatrix} \cos \Omega & \sin \Omega \\ -\sin \Omega & \cos \Omega \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix} \begin{pmatrix} \cos \Omega & -\sin \Omega \\ \sin \Omega & \cos \Omega \end{pmatrix}, \quad (\text{A.12})$$

leading to

$$\sigma_{x_n}^2 = \sigma_x^2 \cos^2 \Omega + \sigma_y^2 \sin^2 \Omega + 2 \cos \Omega \sin \Omega \rho \sigma_x \sigma_y \quad (\text{A.13})$$

$$\sigma_{y_n}^2 = \sigma_x^2 \sin^2 \Omega + \sigma_y^2 \cos^2 \Omega - 2 \cos \Omega \sin \Omega \rho \sigma_x \sigma_y \quad (\text{A.14})$$

$$\rho_n \sigma_{x_n} \sigma_{y_n} = \cos \Omega \sin \Omega (\sigma_y^2 - \sigma_x^2) + (\cos^2 \Omega - \sin^2 \Omega) \rho \sigma_x \sigma_y \quad (\text{A.15})$$

A.4. Convolution product of two independent 2D Gaussians

The errors on the X-ray and the optical source are two Gaussian random variables Z_X and Z_o with variance-covariance matrices

$$V_{Z_X} = \begin{pmatrix} \sigma_{x_X}^2 & \rho_X \sigma_{x_X} \sigma_{y_X} \\ \rho_X \sigma_{x_X} \sigma_{y_X} & \sigma_{y_X}^2 \end{pmatrix}, V_{Z_o} = \begin{pmatrix} \sigma_{x_o}^2 & \rho_o \sigma_{x_o} \sigma_{y_o} \\ \rho_o \sigma_{x_o} \sigma_{y_o} & \sigma_{y_o}^2 \end{pmatrix}. \quad (\text{A.16})$$

The random variable $Z = Z_X + Z_o$ is define by the convolution product of Z_X by Z_o . We know that the sum of 2 Gaussians is a Gaussian with a variance-covariance matrix:

$$V_Z = \begin{pmatrix} \sigma_{x_X+x_o}^2 & \rho \sigma_{x_X+x_o} \sigma_{y_X+y_o} \\ \rho \sigma_{x_X+x_o} \sigma_{y_X+y_o} & \sigma_{y_X+y_o}^2 \end{pmatrix}. \quad (\text{A.17})$$

As Z_X and Z_o are independent, x_X and x_o are independent. Idem for y_X and y_o . We thus have $\text{Cov}(x_X, x_o) = \text{Cov}(y_X, y_o) = 0$. The variance formula $\text{Var}(a + b) = \text{Var}(a) + \text{Var}(b) + 2\text{Cov}(a, b)$ leads to

$$\sigma_{x_X+x_o}^2 = \sigma_{x_X}^2 + \sigma_{x_o}^2 \quad (\text{A.18})$$

$$\sigma_{y_X+y_o}^2 = \sigma_{y_X}^2 + \sigma_{y_o}^2. \quad (\text{A.19})$$

The covariance is given by $\rho \sigma_{x_X+x_o} \sigma_{y_X+y_o} = E\{(x_X + x_o)(y_X + y_o)\} - E\{(x_X + x_o)\}E\{(y_X + y_o)\}$. But as our four 1D Gaussian x_X , x_o , y_X and y_o are centred, their mean is null and thus $E\{(x_X + x_o)\} = E\{x_X\} + E\{x_o\} = 0$ and $E\{(y_X + y_o)\} = E\{y_X\} + E\{y_o\} = 0$. It leads to $\rho \sigma_{x_X+x_o} \sigma_{y_X+y_o} = E\{(x_X y_X)\} + E\{(y_X y_o)\} + E\{(x_X y_o)\} + E\{(y_X x_o)\}$, which is a sum of covariance since our four distributions are centred. We already mentioned that the covariance of x_X and x_o is null. Idem for y_X and y_o . We finally obtain

$$\rho \sigma_{x_X+x_o} \sigma_{y_X+y_o} = \rho_X \sigma_{x_X} \sigma_{y_X} + \rho_o \sigma_{x_o} \sigma_{y_o}. \quad (\text{A.20})$$

Appendix B: Estimate of the local density

In the 1σ ellipse (Eq. (6)), which is the elementary surface, the Poissonian density of sources is $\lambda = \pi \sigma_M \sigma_m f(\alpha_X, \delta_X, m_o)$. The accuracy of the likelihood ratio (Eq: (9)) depends sensitively on the local density estimation $f(\alpha_X, \delta_X, m_o)$. A simple manner to estimate the local density of the SDSS-DR7 catalogue is to use a

k nearest neighbour (knn) averaging. In this case, the estimated densities have a well known error: $\sqrt{k}/\pi d_k^2$, with d_k the distance of the k th neighbour from the X-ray source. However, knn averaging has the impeding property of being non-differentiable. In fact, two candidates with similar magnitudes – or two nearby sources with the same magnitude – could have quite different LR s. The Voronoi tessellation and the use of wavelets also seems too complicated and time-consuming for our simple purpose.

Considering these drawbacks, we preferred to use the sample-point density estimator, which is a kernel smoothing, using the Epanechnikov profile as kernel function and the distance to the 100th nearest neighbours as bandwidth.

Kernel smoothing – also called Parzen window technique or kernel density estimation – can be performed with a fixed or a variable bandwidth. Zhang et al. (2006) provide Markov Chain Monte Carlo (MCMC) algorithms for estimating optimal data-driven fixed bandwidths. These algorithms involve the creation of a density map. In our case we consider for each candidate the density of sources at least as bright as the candidate. We therefore should compute roughly as many density maps as the number of candidates and apply several times the algorithms. Since the MCMC technique in our case is time-consuming, it has been discarded.

We therefore investigated variable bandwidth kernel smoothing techniques. The simplest one is the *balloon* estimator (see Sain 2002). Unfortunately it usually fails to integrate to one. We therefore preferred a *sample-point density estimator* in which the estimate of the density at a point \mathbf{x} of a FOV containing N sources at positions \mathbf{x}_i , $i \in [[1, N]]$ is given by

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^N \frac{1}{h(\mathbf{x}_i)^2} K_E \left(\frac{\|\mathbf{x} - \mathbf{x}_i\|}{h(\mathbf{x}_i)} \right), \quad (\text{B.1})$$

where K is a kernel function and $h(\mathbf{x}_i)$ a bandwidth depending on the position. According to Comaniciu et al. (2001), this estimator “is proved to be almost all the time much better than the fixed bandwidth estimator”. The article gives an usual form of the variable bandwidth $h(\mathbf{x}_i)$:

$$h(\mathbf{x}_i) = h_o \left[\frac{\lambda}{f(\mathbf{x}_i)} \right]^{1/2}, \quad (\text{B.2})$$

where h_o represents a fixed bandwidth, λ a proportionality constant impacting on the smoothness and f a pilot function. It is also mentioned that “the method is insensitive to the fine detail of the pilot estimate”, and “a good initial choice is to take λ as the geometric mean” of the pilot function.

The selected pilot function stems from a knn averaging: $f(\mathbf{x}_i) = \frac{k}{\pi d_k^2}$ where d_k is the distance to the k th nearest neighbour at least as bright as the candidate. It is fast to compute and the discontinuities can be considered as negligible *fine details*. As suggested in Comaniciu et al. (2001), we define λ as the mean value of the pilot function: $\lambda = \frac{N}{\pi R_{fov}^2}$, where N is the total number of sources in the field of view with a magnitude lower than or equal to that of the candidate, and where R_{fov} is the radius of the FOV. We used a fixed bandwidth, which corresponds to the area in which we find, on the average, k sources: $h_o = \sqrt{\frac{k}{N}} R_{fov}$. This leads to $h(\mathbf{x}_i) = d_k$. In order to complete our estimate of the local density we only need to define parameter k , the number of nearest neighbours being at least as bright as the candidate considered.

The smaller is k , the more local is the density, but the larger is the error on the density estimation. Using $k = 100$ leads to a relative error on LR of about 15%, which corresponds to 6.5% for $\log_{10} LR$. In addition, we implemented an algorithm handling border effects.

The selected kernel profile is the 2D Epanechnikov profile K_E :

$$K_E(\tau) = \begin{cases} \frac{2}{\pi}(1 - \tau^2) & \text{if } \tau < 1 \\ 0 & \text{else.} \end{cases} \quad (\text{B.3})$$

This truncated parabola allows computational optimisations thanks to its finite extent. Moreover, since it minimises the mean integrated square error (Comaniciu & Meer 2002), this kernel maximises the quality of the density reconstruction.

Appendix C: Some details on the likelihood ratio

We consider the case where an X-ray source has only one candidate and formulate two hypotheses:

H_{cp} : the candidate is the counterpart, it is not random.
 H_{spur} : the candidate is a random source belonging to a Poissonian distribution of sources of density λ .

The union of these two hypotheses gives the set of all possibilities: $P(H_{cp}) + P(H_{spur}) = 1$, i.e. $H_{cp} = \bar{H}_{spur}$. The probability of finding a candidate at a distance r of the X-ray source can therefore be written as:

$$P(r) = P(r|H_{cp})P(H_{cp}) + P(r|H_{spur})P(H_{spur}). \quad (\text{C.1})$$

From conditional probabilities

$$P(H_{cp}|r) = \frac{P(H_{cp} \cap r)}{P(r)} \quad (\text{C.2})$$

$$P(r|H_{cp}) = \frac{P(H_{cp} \cap r)}{P(H_{cp})} \quad (\text{C.3})$$

we deduce

$$P(H_{cp}|r) = \frac{P(r|H_{cp})P(H_{cp})}{P(r)}, \quad (\text{C.4})$$

which can be written (Eq. (C.1)) as:

$$P(H_{cp}|r) = \frac{P(r|H_{cp})P(H_{cp})}{P(r|H_{cp})P(H_{cp}) + P(r|H_{spur})P(H_{spur})}. \quad (\text{C.5})$$

We transform that equation into:

$$P(H_{cp}|r) = \frac{1}{1 + \frac{1}{\frac{P(H_{cp})}{P(H_{spur})} LR}}, \quad (\text{C.6})$$

to exhibit the likelihood ratio

$$LR(r) = \frac{P(r|H_{cp})}{P(r|H_{spur})}. \quad (\text{C.7})$$

The remaining term $\frac{P(H_{cp})}{P(H_{spur})}$ has to be estimated to compute $P(H_{cp}|r)$, the probability that the candidate is a real counterpart. Considering the estimate of the number of real counterparts divided by the number of spurious associations, it leads to the $\theta/(1 - \theta)$ factor of De Ruiter et al. (1977).

If we add a prior knowledge on the magnitude, the formula (C.6) is unchanged except that it now expresses $P(H_{cp}|r \cap m)$ and the LR becomes

$$LR(r \cap m) = \frac{P(r \cap m|H_{cp})}{P(r \cap m|H_{spur})} = \frac{P(r|H_{cp})P(m|H_{cp})}{P(r|H_{spur})P(m|H_{spur})}. \quad (\text{C.8})$$

The term $P(r \cap m|H_{spur}) = P(r|H_{spur})P(m|H_{cp})$ expresses the probability of having a spurious source of magnitude m . It is the Poissonian local distribution of sources of magnitude m . The term $P(r|H_{cp})P(m|H_{cp})$ is the classical 2D Gaussian distribution times the probability of having, among all real counterparts, a counterpart of magnitude m . $P(m|H_{cp})$ can be seen as the distribution of counterpart sources according to their magnitudes and correspond to the $q(m)$ factor in Brusa et al. (2007) and of $q(m, c)$ in Sutherland & Saunders (1992) if we ignore object types.

Appendix D: Computation of the spurious LR distribution

Let us first consider the case of one X-ray and one optical source of magnitude m in a FOV. The probability of observing the random X-ray source at a given point (α_X, δ_X) of the FOV is given by the distribution $g_X(\alpha_X, \delta_X)$ of the X-ray sources in the FOV. The chance for the optical source of magnitude m to be randomly associated with the X-ray source is given by the integral of the distribution of optical sources at least as bright as m over the surface of the convolution ellipse $S_{x,o} = k_\gamma^2 \pi \sigma_M \sigma_m$. The probability that the association has a LR in a bin ΔLR is then given by the integral of the distribution of optical sources at least as bright as m over the surface the bin occupies inside the convolution ellipse $S_{x,o}$: $S_{x,o}(\Delta LR)$. Finally, the sum of these probabilities for all X-ray and archival sources provides an estimate of the number of spurious associations by LR bin. Considering now the case of N X-ray and M optical sources in the FOV leads to the formula of $N_{spur}(\Delta LR)$

$$\sum_{x=1}^N \iint_{S_{FOV}} g_X(\alpha_X, \delta_X) \left(\sum_{o=1}^M \iint_{S_{x,o}(\Delta LR)} g_o(\alpha_o, \delta_o, m) d\alpha_o d\delta_o \right) d\alpha_X d\delta_X, \quad (\text{D.1})$$

where $S_{x,o}(\Delta LR)$ depends on the X-ray source, the optical source, and the position of the X-ray source and can therefore be written as $S_{x,o}(\Delta LR, X, o, \alpha_X, \delta_X)$.

A possible manner to make the problem simpler is to assume that $g_o(\alpha_o, \delta_o, m)$ is roughly constant inside the surface $S_{x,o}(\Delta LR)$. Eq. (D.1) can then be written as

$$\iint_{S_{FOV}} g_X(\alpha_X, \delta_X) \left(\sum_{x=1}^N \sum_{o=1}^M S_{x,o}(\Delta LR) g_o(\alpha_X, \delta_X, m) \right) d\alpha_X d\delta_X. \quad (\text{D.2})$$

In order to estimate $N_{spur}(\Delta LR)$ we have to compute a complete distribution map of the X-ray sources and as many distribution maps as the number of distinct optical source magnitudes. Since such a computation is potentially time-consuming, we may suppose that optical sources are uniformly distributed in the FOV and have as local density the mean value of their local densities. This less rigorous approach leads to the simple formula

$$N_{spur}(\Delta LR) = \sum_{x=1}^N \sum_{o=1}^M S_{x,o}(\Delta LR) / S_{FOV}. \quad (\text{D.3})$$

The quantity $S_{x,o}(\Delta LR)$ can be computed as follows: the radius corresponding to a given LR is easily derived from Eq. (9):

$$r^2(LR) = \begin{cases} 0 & \text{if } LR \geq LR_{\max} = \frac{1}{2\lambda} \\ \ln \frac{1}{(2\lambda LR)^2} & \text{if } LR_{\min} < LR < LR_{\max} \\ k_\gamma^2 & \text{if } LR \leq LR_{\min} = \frac{1}{2\lambda} e^{\frac{1}{2} k_\gamma^2}. \end{cases} \quad (\text{D.4})$$

Thus for a bin $\Delta LR = LR_2 - LR_1$, $LR_2 > LR_1$, if we have $LR_1 \leq LR_{\max}$ and $LR_2 \geq LR_{\min}$, the surface of the bin of LR in the ellipse is expressed by

$$S_{x,o}(\Delta LR) = 2\pi\sigma_M\sigma_m \ln \frac{\min(\max(LR_2, LR_{\min}), LR_{\max})}{\max(\min(LR_{\max}, LR_1), LR_{\min})}. \quad (\text{D.5})$$

In principle, we should remove the optical sources which are the real counterparts since they are not randomly distributed. However, their fraction can be neglected at low LR , and at very high LR their influence on the final probability of identification is insignificant. For real data, the source distribution is not completely Poissonian because two sources cannot be infinitely close due to the PSF, the resolution of the instrument, and the extraction algorithm. There is therefore a small surface around each counterpart which cannot contain another source. Neglecting this can lead to slightly overestimate the rate of spurious associations.

Appendix E: Implementation

All software was written in Java. 2XMMi sources were retrieved using direct SQL queries – through a Java DataBase Connection – on the XCat-DB. SDSS DR7 sources were collected from the VO thanks to the ConeSearch protocol, the correlation has thus been performed independently for each XMM observation.

Local density estimations required a lot of k -nearest-neighbours search. Moreover, for each candidate we are only interested in sources with a magnitude smaller than or equal to that of the candidate. In order to quickly perform such knn queries, SDSS DR7 sources have been stored in quadtrees especially designed for that purpose: each leaf of the tree stores the greatest magnitude among those of the sources held in its sub-tree; knn queries contain an extra parameter which is the magnitude of the candidate; in addition to the distance criterion, we thus easily add a magnitude criterion to stop the search in a sub-tree.

A quadtree scheme was preferred to a kd-tree because it is fastest to build and easier to update. It is then possible to build the quadtree during the parsing of the VOTable returned by the ConeSearch query.

References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
 Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297
 Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539
 Antonucci, R. 1993, ARA&A, 31, 473
 Barcons, X., Carrera, F. J., Watson, M. G., et al. 2002, A&A, 382, 522
 Barcons, X., Carrera, F. J., Ceballos, M. T., et al. 2007, A&A, 476, 1191
 Bochanski, J. J., West, A. A., Hawley, S. L., & Covey, K. R. 2007, AJ, 133, 531
 Boller, T., Bertoldi, F., Dennefeld, M., & Voges, W. 1998, A&AS, 129, 87
 Brusa, M., Zamorani, G., Comastri, A., et al. 2007, ApJS, 172, 353
 Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301
 Caccianiga, A., Severgnini, P., Braito, V., et al. 2004, A&A, 416, 901
 Caccianiga, A., Severgnini, P., Della Ceca, R., et al. 2008, A&A, 477, 735
 Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, A&A, 497, 635
 Carrera, F. J., Ebrero, J., Mateos, S., et al. 2007, A&A, 469, 27

- Comaniciu, D., & Meer, P. 2002, *IEEE Trans. Pattern Anal. Mach. Intell.*, 24, 603
- Comaniciu, D., Ramesh, V., & Meer, P. 2001, in *Computer Vision, ICCV 2001, Proc. Eighth IEEE Int. Conf.*, 1, 438
- Covey, K. R., Ivezić, Ž., Schlegel, D., et al. 2007, *AJ*, 134, 2398
- Covey, K. R., Agüeros, M. A., Green, P. J., et al. 2008, *ApJS*, 178, 339
- Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, in *The IRSA 2MASS All-Sky Point Source Catalog, NASA/IPAC Infrared Science Archive*, <http://irsa.ipac.caltech.edu/applications/Gator/>
- De Ruiter, H. R., Arp, H. C., & Willis, A. G. 1977, *A&AS*, 28, 211
- Della Ceca, R., Maccacaro, T., Caccianiga, A., et al. 2004, *A&A*, 428, 383
- Della Ceca, R., Caccianiga, A., Severgnini, P., et al. 2008, *A&A*, 487, 119
- den Herder, J. W., Brinkman, A. C., Kahn, S. M., et al. 2001, *A&A*, 365, L7
- Downes, R. A., Webbink, R. F., Shara, M. M., et al. 2001, *PASP*, 113, 764
- Gilli, R., Comastri, A., & Hasinger, G. 2007, *A&A*, 463, 79
- Girardi, L., Grebel, E. K., Odenkirchen, M., & Chiosi, C. 2004, *A&A*, 422, 205
- Güdel, M., & Nazé, Y. 2009, *A&ARv*, 17, 309
- Guedel, M., Guinan, E. F., & Skinner, S. L. 1997, *ApJ*, 483, 947
- Guillout, P., Klutsch, A., Frasca, A., et al. 2009, *A&A*, 504, 829
- Hands, A. D. P., Warwick, R. S., Watson, M. G., & Helfand, D. J. 2004, *MNRAS*, 351, 31
- Heckman, T. M., Ptak, A., Hornschemeier, A., & Kauffmann, G. 2005, *ApJ*, 634, 161
- Hérent, O., Motch, C., & Guillout, P. 2006, in *The X-ray Universe 2005*, ed. A. Wilson, ESA SP-604, 91
- Jansen, F., Lumb, D., Altieri, B., et al. 2001, *A&A*, 365, L1
- Krumpe, M., Lamer, G., Corral, A., et al. 2008, *A&A*, 483, 415
- Loaring, N. S., Dwelly, T., Page, M. J., et al. 2005, *MNRAS*, 362, 1371
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2010, *ApJS*, 187, 560
- Mainieri, V., Bergeron, J., Hasinger, G., et al. 2002, *A&A*, 393, 425
- Mason, K. O., Breeveld, A., Much, R., et al. 2001, *A&A*, 365, L36
- Michel, L., Motch, C., Nguyen Ngoc, H., & Pineau, F. X. 2009, in *ASP Conf. Ser.* 411, ed. D. A. Bohlender, D. Durand, & P. Dowler, 563
- Motch, C. 2005, in *5 years of Science with XMM-Newton*, ed. U. G. Briel, S. Sembay, & A. Read, 29
- Motch, C. 2006, *The X-ray Universe 2005*, ed. A. Wilson, in ESA SP-604, 383
- Motch, C., Guillout, P., Haberl, F., et al. 1997, *A&A*, 318, 111
- Motch, C., Hérent, O., & Guillout, P. 2003, *Astron. Nachr.*, 324, 61
- Motch, C., Michel, L., & Pineau, F. X. 2007, in *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell, *ASP Conf. Ser.*, 376, 699
- Motch, C., Warwick, R., Cropper, M. S., et al. 2010, *A&A*, 523, A92
- Oyabu, S., Yun, M. S., Murayama, T., et al. 2005, *AJ*, 130, 2019
- Perola, G. C., Puccetti, S., Fiore, F., et al. 2004, *A&A*, 421, 491
- Pineau, F.-X., Derriere, S., Michel, L., & Motch, C. 2008, in *AIP Conf. Ser.*, 1082, 15
- Pineau, F., Derriere, S., Michel, L., & Motch, C. 2009, in *ASP Conf. Ser.*, ed. D. A. Bohlender, D. Durand, & P. Dowler, 411, 259
- Ptak, A., Zakamska, N. L., Strauss, M. A., et al. 2006, *ApJ*, 637, 147
- Reyes, R., Zakamska, N. L., Strauss, M. A., et al. 2008, *AJ*, 136, 2373
- Richards, G. T., Nichol, R. C., Gray, A. G., et al. 2004, *ApJS*, 155, 257
- Ritter, H., & Kolb, U. 2003, *A&A*, 404, 301
- Rogel, A. B., Lugger, P. M., Cohn, H. N., et al. 2006, *ApJS*, 163, 160
- Rutledge, R. E., Brunner, R. J., Prince, T. A., & Lonsdale, C. 2000, *ApJS*, 131, 335
- Sain, S. R. 2002, *Comput. Stat. Data Anal.*, 39, 165
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schmidt, M., Hasinger, G., Gunn, J., et al. 1998, *A&A*, 329, 495
- Schwöpe, A., Hasinger, G., Lehmann, I., et al. 2000, *Astron. Nachr.*, 321, 1
- Severgnini, P., Caccianiga, A., Braitto, V., et al. 2006, *A&A*, 451, 859
- Sinnott, R. W. 1984, *S&T*, 68, 158
- Stephen, J. B., Bassani, L., Molina, M., et al. 2005, *A&A*, 432, L49
- Stoeck, J. T., Morris, S. L., Gioia, I. M., et al. 1991, *ApJS*, 76, 813
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, 124, 1810
- Strüder, L., Briel, U., Dennerl, K., et al. 2001, *A&A*, 365, L18
- Sung, H., Bessell, M. S., & Sana, H. 2008, *J. Korean Astron. Soc.*, 41, 1
- Sutherland, W., & Saunders, W. 1992, *MNRAS*, 259, 413
- Tedds, J. A., Page, M. J., & Survey Science Centre, X. 2006, in *The X-ray Universe 2005*, ed. A. Wilson, ESA SP-604, 843
- Treister, E., Urry, C. M., & Virani, S. 2009, *ApJ*, 696, 110
- Turner, M. J. L., Abbey, A., Arnaud, M., et al. 2001, *A&A*, 365, L27
- Tyson, J. A. 2002, in *SPIE Conf. Ser.* 4836, ed. J. A. Tyson, & S. Wolff, 10
- Ueda, Y., Watson, M. G., Stewart, I. M., et al. 2008, *ApJS*, 179, 124
- Vaiana, G. S., Cassinelli, J. P., Fabbiano, G., et al. 1981, *ApJ*, 245, 163
- Véron-Cetty, M.-P., & Véron, P. 2006, *A&A*, 455, 773
- Vignali, C., Alexander, D. M., & Comastri, A. 2006, *MNRAS*, 373, 321
- Vignali, C., Alexander, D. M., Gilli, R., & Pozzi, F. 2010, *MNRAS*, 404, 48
- Wang, J., Protopapas, P., Chen, W., et al. 2010, *AJ*, 139, 2003
- Watson, M. G., Auguères, J.-L., Ballet, J., et al. 2001, *A&A*, 365, L51
- Watson, M. G., Schröder, A. C., Fyfe, D., et al. 2009, *A&A*, 493, 339
- Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P., & O'Dell, S. L. 2000, in *SPIE Conf. Ser.* 4012, ed. J. E. Truemper, & B. Aschenbach, 2
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, 154, 1
- White, R. L. 2008, in *AIP Conf. Ser.* 1082, ed. C. A. L. Bailer-Jones, 37
- Wolstencroft, R. D., Savage, A., Clowes, R. G., et al. 1986, *MNRAS*, 223, 279
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2003, *AJ*, 126, 2125
- Zhang, X., King, M. L., & Hyndman, R. J. 2006, *Comput. Stat. & Data Anal.*, 50, 3009
- Zickgraf, F.-J., Thiering, I., Krautter, J., et al. 1997, *A&AS*, 123, 103