Mem. S.A.It. Suppl. Vol. 17, 106

Memorie della
Supplementi

Identification of (high-redshift) AGN with WFXT: lessons from COSMOS and CDFS

M. Brusa[1], R. Gilli[2], F. Civano[3], A. Comastri[2], R. Fiore[4], and C. Vignali[5]

[1] Max Planck Institut für Extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching by München, Germany e-mail: marcella@mpe.mpg.de
[2] Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Bologna, Via Ranzani 1, I-40127 Bologna, Italy
[3] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138
[4] Istituto Nazionale di Astrofisica – Osservatorio Astronomico di Monteporzio Catone, Via Frascati 33, I-00044 Monte Porzio Catone, Italy
[5] Dipartimento di Astronomia – Università di Bologna, Via Ranzani 1, I-40127 Bologna, Italy

Abstract. The Wide Field X–ray Telescope (WFXT) will provide tens of millions of AGN, with more than 4×10^5 expected at z> 3. Here we review the issues present in the identification of (large) samples of faint and high-redshift X–ray sources, and describe a statistical, powerful tool that can be applied to WFXT catalogs. The depth of associated optical and near infrared catalogs, needed for a reliable and as much complete as possible identification, are also discussed, along with the combined synergies with existing or planned facilities.

Key words. Galaxies: active – X-rays: Active Galactic Nuclei – galaxies: high-redshift

1. Scientific drivers

One of the main aims in extragalactic astronomy for the next decade is the study of the co-evolution of galaxies and the Super Massive Black Holes (SMBH) residing in their centre, out to the very first epochs of galaxy formation. In this respect, deep and sensitive X–ray observations will be the *unique* instrument to reveal the accretion light from SMBH in galactic nuclei at high-z, which are often invisible at longer wavelengths because of intergalactic absorption and dilution by the host galaxy.

The study of Active Galactic Nuclei (AGN) demography at z> 3 is one of the key science drivers for the Wide Field X–ray Telescope (WFXT, e.g. Forman et al. 2010). In the past decade, the characterization of the early phase of SMBH growth has been limited to the study of optically selected QSOs detected mostly in the SDSS survey, i.e. sampling only the unobscured and most luminous tail of the AGN population. Deep and medium deep *Chandra* and XMM-*Newton* surveys have allowed the study of X–ray selected QSOs up to relatively high redshifts, z∼ 3 – 4. At higher redshifts, present X–ray surveys are highly incomplete and strongly limited by the small area sampled. As an example, there are only a few X–ray selected QSO with confirmed spectroscopic redshifts at z> 5 (see Barger et al. 2005) Moreover, the extrapolations of the X–ray lu-

Send offprint requests to: M. Brusa

minosity function (LF) as obtained combining various XMM and Chandra surveys differ by up an order of magnitude (see Figure 1, and reference therein, adapted from Brusa et al. 2010). As a comparison, the number of optically selected QSOs revealed up to z∼ 6 is approaching 50, i.e. large enough to determine their LF which encodes the information about the history of SMBH build up and the integrated flux of UV ionizing radiation (e.g. Fan et al. 2006; Jiang et al. 2009; Willott et al. 2010, and reference therein). An unbiased search of X–ray selected z∼ 5 – 6 QSOs would require to survey several hundreds of square degrees to a depth of the order of ∼ 10^{-15} erg cm^{-2} s^{-1} and thus beyond the capabilities of current X–ray telescopes. WFXT will offer the unique opportunity to explore the high-redshift universe, providing *about two order of magnitudes* larger samples with respect to the current SDSS samples (∼ 2000 z> 6 AGN vs. ∼ 50), opening a completely new, unexplored window for LF analysis. (see Gilli et al. 2010, this volume, for a full description of the high-redshift AGN demography with WFXT).

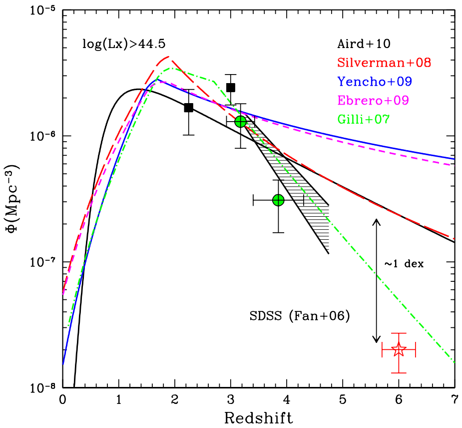

Fig. 1. The number density of $\log L_{2-10keV} > 44.5$ erg s^{-1} X–ray selected AGN vs. redshift as obtained from published LF, as labeled (Gilli, Comastri & Hasinger 2007; Silverman et al. 2008; Ebrero et al. 2009; Yencho et al. 2009; Aird et al. 2010) along with datapoints from the XMM–COSMOS surveys (green circles, Brusa et al. 2009a) and from Aird et al. (2010, black squares). The red star at $z = 6$ represents a conservative estimate of the z∼ 6 QSO space density computed from the optical one assuming no evolution of the $\alpha_{\rm ox}$. For comparison (black shaded area) we plot the results for very bright QSOs ($M_{1450} < -27$, Fan et al. 2006), rescaled by an arbitrary factor.

2. Identification issues

The identification of the correct counterparts of both obscured and unobscured AGN is the first, crucial step for a full characterization of the physical and evolutionary properties of the entire population. At high redshifts, this process is further complicated by the fact that 1) z> 3 sources constitute only a tiny fraction (∼ 1%) of the entire X–ray population (< 0.1% for z> 6 sources) and 2) these objects are usually faint in the optical band, because the emission would be strongly reduced by cosmological dimming, and/or, for obscured sources, the intrinsic AGN emission is absorbed by the surrounding material. As a result, the probability of finding by chance a galaxy of R> 24 in the X–ray error box is not negligible even with Chandra given the high surface density of background galaxies (see extensive discussions in, e.g., Luo et al. 2010). The identification process is made easier by using deep near infrared images given that AGN are strong IR emitters and the K-band flux is more tightly correlated with the X-ray flux than the optical (obscured) one (see Brusa et al. 2009b).

2.1. The likelihood ratio technique

A statistical, powerful method extensively exploited in deep XMM-Newton and Chandra surveys in the past years to look for the correct counterparts of X–ray sources is the "likelihood ratio" (*LR*) technique (Sutherland & Saunders 1992; Brusa et al. 2005). The method calculates the probability that a source is the correct association by weighting the information on the X–ray to optical distance, the surface density of (possible) false coincidence background objects and the brightness of the chosen counterpart:

$$LR = \frac{q(m)f(r)}{n(m)}$$

The object with the highest LR value[1] (above a certain threshold; see Sutherland & Saunders 1992 for details) is the most likely counterpart; when two or more sources have comparable LR values, a unique identification is not possible and both sources have a similar probability of being the correct identification ("ambiguous" sources). Using catalogs extracted from different bands (e.g., optical and infrared) may lead to different choices of the correct counterparts, and this information should be taken into account, too. In the following we will show the potentiality (and the challenges) on the use of the LR technique applied to WFXT data and the multiwavelength datasets available. We will make the case separately for the Wide ($F_{0.5-2} \gtrsim 3 \times 10^{-15}$ erg cm^{-2} s^{-1}), Medium ($F_{0.5-2} \gtrsim 5 \times 10^{-16}$ erg cm^{-2} s^{-1}), and Deep ($F_{0.5-2} \gtrsim 3 \times 10^{-17}$ erg cm^{-2} s^{-1}) parts of the WFXT survey (Rosati et al. 2010, this volume), based on the experience developed in the framework of the XMM-COSMOS (Hasinger et al. 2007), C-COSMOS (Elvis et al. 2009) and CDFS (Luo et al. 2008) surveys, where *multiwavelength* catalogs (e.g. optical to mid-infrared) resulted crucial to keep the fraction of ambiguous or false identification at minimim.

2.2. Wide and Medium survey: COSMOS lessons

To quantify the expected efficiency of the LR technique on the sources detected in the WFXT Wide survey, we first limited the XMM-COSMOS sample (Cappelluti et al. 2009) at fluxes larger than the expected limiting flux of the WFXT Wide survey ($F_{0.5-2keV} > 3 \times 10^{-15}$ erg cm^{-2} s^{-1}) and comparable to those expected for the eROSITA deep survey (Predehl et al. 2010, see also Cappelluti et al. 2010, this volume). Then we looked at the breakdown of the combined optical and IR identifications from the LR technique: 95% of the sources have been provided "secure" associations, while the remaining 5% show ambiguous counterparts in the Brusa et al. (2010) catalog. The reliability of the method has been tested a posteriori using Chandra, and turned out to be 99.6%: only one of the 245 unique/reliable XMM-COSMOS sources at fluxes larger than the WFXT Wide survey resulted associated to the wrong optical counterpart. Moreover, the statistical properties (such as redshifts, magnitudes, colors etc.) of the primary and secondary counterpart within the ambiguous sources are almost *indistinguishable*, and therefore the choice of the counterpart among the two does not in principle affect the characterization of the full X–ray population.

The WFXT Medium survey has been desgigned to cover ~ 3000 deg^2 at fluxes of the order of $\sim 5\times10^{-16}$ erg cm^{-2} s^{-1}, i.e. comparable to the depth reached by the C-COSMOS survey (Elvis et al. 2009). Following a procedure similar to that applied to XMM-COSMOS data (see above), and thanks to the smaller (< 1") angular resolution of *Chandra* with respect to XMM-*Newton*, Civano et al. (2010) were able to provide secure associations for more than 95% of the sources detected above the expected WFXT Medium survey limit. The fraction of ambiguous sources in this sample is reduced to $\sim 2\%$, only 1.4% of the X–ray sources are not identified.

Taking into account that the WFXT positional accuracy is expected to be better than that of XMM-*Newton* (HEW=5" – 10" for WFXT vs. HEW $\sim$ 15" for XMM-*Newton*), and only slightly worse than the *Chandra* one (HEW $\sim$ 2" when averaged across the FOV), we can safely conclude that counterpart identification would not be an issue for the WFXT Wide and Medium surveys, provided that the depth of the optical and IR ancillary data is enough to match the X–ray fluxes (see Section 2.4).

2.3. Deep survey: CDFS lessons

The WFXT deep survey ($F_{0.5-2keV} > 3 \times 10^{-17}$ erg cm^{-2} s^{-1} in the soft X–ray band) would be almost as deep as the CDFS 2Ms survey

[1] *q(m)* is the expected probability distribution, as a function of magnitude, of the true counterparts, *f(r)* is the probability distribution function of the positional errors of the X–ray sources assumed to be a two–dimensional Gaussian, and *n(m)* is the surface density of background objects with magnitude *m*.

Table 1. Optical and IR ideal coverage depth for WFXT AGN surveys

Survey	$F^{lim}_{0.5-2}$ cgs	I	K
Wide	4×10^{-15}	23.0	21.5
Medium	5×10^{-16}	25.0	23.0
Deep	3×10^{-17}	25.5	23.5

(Luo et al. 2008), over an area that is a factor of ~ 1000 larger. The detailed identification analyses for the 2Ms CDFS sources (Luo et al. 2010), implementing likelihood ratio matching accross five bands (R, z, K, 3.6μm and 1.4 GHz) have shown the power of this approach while also quantifying the significant challenges in source identification at faint magnitudes (R> 25). Indeed, it was possible to provide identifications for 96% of the X–ray sources; among them, 90% have been classified as unique/secure, and 10% as ambiguous. At these deep X–ray fluxes, the statistical properties of primary and secondary counterparts between the ambiguous sources are often *different*. Moreover, 4% of the sources, despite the excellent quality and depth of the multiwavelength information available, were not associated to any counterpart, i.e. the correct counterpart is most likely fainter than the image depth. This exercise shows that the identification of the faintest among the WFXT counterparts in the deep survey may be challenging; In this respect, the HEW goal of 5” is a crucial requirement to keep at acceptable values the (already not negligible) identifications issues and to fully characterize the multiwavelength properties of the X–ray sources at the highest redshifts.

2.4. Depth of optical infrared images

The power of the LR technique described in the previous subsections is strongly related to the depth of the optical and infrared images and catalogs that will be used to identify the X-ray sources. The challenge will be to provide a homogeneous and (enough) deep coverage for the different WFXT surveys. At the limiting flux of the WFXT wide survey an optical coverage to I~ 23 and K~ 21.5 would be enough to identify ~ 90% of the X–ray sources (see Figure 2, upper panels, and Table 1), but this should be on the *entire* surveyed area. At the time WFXT will be launched, PanSTARRS[2] will have surveyed ~ 30.000 deg^2 to I~ 24.2, and will provide imaging in at least 5 bands, needed to characterize the SED of the X–ray sources and isolate high-z candidates (see next Section). On a longer timescale, Euclid[3] will cover the entire extragalactic sky in the IR down to H~ 24 (roughly corresponding to K~ 23), and will provide also spectra. The LOw Frequency ARray (LOFAR, Morganti et al. 2010), that will survey the northern sky down to a flux of 0.8 mJy at 120 MHz (see Fig. 2 in Morganti et al. 2010), may be crucial to correctly identify radio emitters X–ray sources (radio AGN and starbursts).

PanSTARRS will also provide identification for a substantial fraction (> 50%) of the sources detected in the WFXT Medium and Deep surveys. In order to identify a fraction as large as 90% of the sources in these surveys, a coverage in the optical and near–infrared down to I~ 25.5 and K~ 23.5 is needed (see Figure 2, middle and lower panels, and Table 1). LSST is a proposed facility expected to cover the southern sky down to I~ 27 (Abell et al. 2010); similarly to PanSTARRS, LSST will also provide multiband photometry at a depth comparable to the I-band limit. The coordination with present and next generation facilities is mandatory, in order to choose the areas for the deep surveys which maximize the availability of the deepest multiwavelength coverage, in particular: JWST[4], the PanSTARRS deep survey (I~ 28 over 28 deg^2), the LSST deep survey (I~ 28 on a few hundreds deg^2), Euclid (K~ 25.5 on 40 deg^2), the VISTA VIDEO survey (down to K=23.5 over 15 deg^2).

[2] http://pan-starrs.ifa.hawaii.edu
[3] http://sci.esa.int/euclid
[4] http://www.jwst.nasa.gov

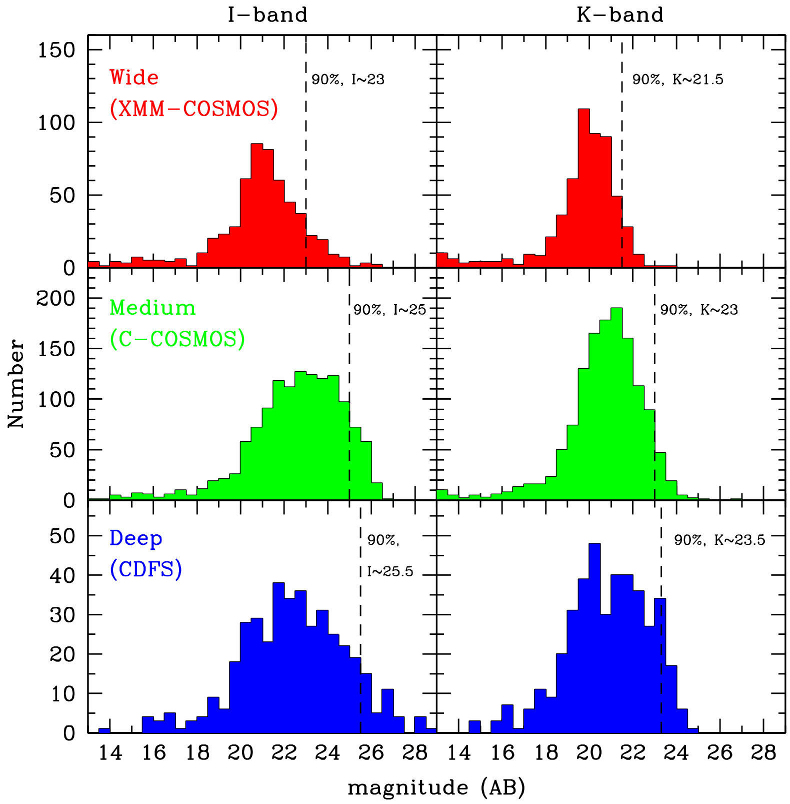

Fig. 2. I-band (left panels) and K-band (right panel) magnitude distributions expected in the three different WFXT surveys (Wide, Medium and Deep, from top to bottom). The expected magnitude distributions have been extracted from the XMM-COSMOS (Brusa et al. 2010), C-COSMOS (Civano et al. 2010), and CDFS (Luo et al. 2010) samples limited to fluxes $F_{0.5-2} > 3 \times 10^{-15}$ erg cm^{-2} s^{-1}, $F_{0.5-2} > 5 \times 10^{-16}$ erg cm^{-2} s^{-1} and $F_{0.5-2} > 3 \times 10^{-17}$ erg cm^{-2} s^{-1} in order to match the Wide, Medium and Deep limiting fluxes, respectively. The dashed lines mark the magnitudes at which most of the sources (90%) are identified.

3. Selecting z>3 (or z>6) AGN

Photometric redshifts of X–ray selected faint sources (R= 24 – 27) are essential for enabling science analyses and planning deep spectroscopy, and resulted *crucial* in isolating the high-z population in, e.g., XMM-COSMOS and CDFS. A similar, detailed source characterization requiring *multiband* imaging may be feasible only for small samples of WFXT sources. In this context, key resources will be again the upcoming LSST, PanSTARRS, Euclid, and (as far as spectroscopy for the Wide survey is concerned) the SDSSIII-BOSS[5] project. However, the high-z population shows on average fainter magnitudes than the overall X–ray source population (see Figure 3 in Brusa et al. 2009a), and therefore may remain among the unidentified population, if deep enough optical and near infrared coverage is not provided over the full area Another possibility is to search for X–ray counterparts on preselected high-z QSO on the basis of optical colours and/or dropouts techniques (e.g. Casey et al. 2008; Steidel et al. 2003), extended including the

[5] http://www.sdss3.org/boss.php

near-infrared bands in order to sample the z> 6 population (e.g. Willott et al. 2010). In this respect, the unprecedent combination of depth and area of WFXT will result in a much better characterization of the physical properties (such as bolometric luminosity and accretion rate) of the first accreting supermassive black holes. Moreover, z> 6 color selections suffer from significant contamination stellar objects (brown dwarfs are overwhelmingly more abundant and the spectroscopy success rate for z> 6 QSOs is only ∼ 20%). The complete SED characterization from NIR to X–ray will be able to resolve issues on contamination and completeness. For a non negligible fraction of the high-z candidates (a few out of a few hundreds, see also Matt et al. 2010, this volume), redshifts may be directly measured from the FeKα line (see examples in Civano, Comastri & Brusa 2005).

4. Conclusions

- WFXT will provide orders of magnitudes *larger* samples of high-redshift (z> 6) AGN compared to current (e.g. SDSS) optical surveys;
- the counterpart identification for WFXT sources selected in the Wide survey will be relatively easy, *if* synergies with present and future large area / all sky facilities (e.g. PanSTARRs, LSST, Euclid) are pursued;
- the secure identification of the counterparts detected in the WFXT Medium and Deep surveys would greatly benefit of the *smallest possible angular resolution* (the 5" HEW goal is really auspicable) and should heavily rely in the coordination with the future optical and NIR deep survey area (e.g. LSST, JWST);
- multiwavelength information is *mandatory* in order to get the redshift and the physical properties of the high-z AGN in the WFXT surveys.

Acknowledgements. We gratefully acknowledge the essential contribution from the COSMOS and CDFS teams, and in particular Bin Luo. RG acknoweledges support from the ASI grant I/088/06/00.

References

Abell et al., 2010, LSST White Book v.2.0, arXiv:0912.0201
Aird, J., et al., 2010, MNRAS, 401, 2531
Barger, A.J., et al., 2005, AJ, 129, 578
Brusa, M., et al., 2005, A&A, 432, 69
Brusa, M., et al., 2009a, ApJ, 693, 8
Brusa, M., et al., 2009b, A&A, 507, 1277
Brusa, M., et al., 2010, ApJ, 716, 348
Cappelluti, N., et al., 2009, A&A, 497, 635
Casey, C., et al., 2008, ApJS, 177, 131
Civano, F., Comastri, A., & Brusa, M., 2005, MNRAS, 358, 693
Civano, F., et al., 2010, ApJ submitted
Ebrero, J., et al. 2009, A&A, 493, 55
Elvis, M., et al., 2009, ApJS, 184, 158
Fan, X., et al. 2006, AJ, 132, 117
Fiore, F., et al., 2010, Proceedings of the conference "X-ray Astronomy 2009", arXiv:1002.3538
Forman, W., et al., 2010, Proceedings of the conference "X-ray Astronomy 2009", arXiv:
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Hasinger, G., et al., 2007, ApJS, 172, 29
Jiang, et al., 2009, AJ 138, 305
Luo, B., et al., 2008, ApJS, 179, 19
Luo, B., et al., 2010, ApJS, 187, 560
Morganti, R., et al., 2010, proceedings of "Panoramic Radio Astronomy: Wide-field 1-2 GHz research on galaxy evolution", arXiv:1001.2384
Predehl, P., et al., 2010, Proceedings of the conference "X-ray Astronomy 2009", arXiv:1001.2502
Silverman, J.D., et al., 2008, ApJ, 679, 118
Steidel, C., et al., 2003, ApJ, 592, 728
Sutherland, W. & Saunders, W. 1992, MNRAS, 259, 413
Yencho, B., et al., 2009, ApJ, 698, 380
Willott, C., et al. 2010, AJ, 139, 906