XMM-LSS
INTERNAL
REPORT n°3-Mi

May 4, 2007

Further progress towards the XMDS full catalogue

A status report

L.Chiappetti[1]

INAF, IASF Milano, via Bassini 15, I-20133 Milano, Italy

Abstract. I report on the work I have done during 2006 and first quarter of 2007 to improve the catalogue of identified sources beyond the original XMDS/VVDS 4 σ catalogue and the subsequent version used for the XMDS/VVDS 3 σ hard sample. In this work I used all database tables available to me in the Milan database up to February 2007, and attempted to automatize the procedure even more than in the previous release, and to use only objective criteria for identification ranking. We intend to use this catalogue to extract subsets for further work since now on.

Key words: LSS – XMDS –

1. Introduction

We originally published in Chiappetti et al. (2005) (hereafter Paper I) a catalogue of 286 tentative identifications for the X-ray sources detected in the XMDS fields at a significance above 4σ, and falling in the area covered by the VVDS survey (Le Fèvre et al., 2004),based on an highly manual procedure, described, together with its input tables, in section 6 of Paper I.

A further working catalogue using all (GO) XMDS sources (1147), inside and outside the VVDS area, using additional data tables which became available in the meantime (e.g. CFHTLS and SWIRE), and automatizing the procedure as much as possible, was produced as described in Chiappetti et al. (2006a), hereafter Report I, and has been used as a basis for the study of the AGN VVDS 3σ hard sample (136 sources) reported by Tajer et al. (2007) and Polletta et al. (2007).

Here we present a new release of a working catalogue for 1168 XMDS sources, which is the ultimate number for the XMDS area. With respect to Report I : *(a)* we included new X-ray sources in the reobserved field G12bis (ESA obs id 0404960501, reobserved in AO-5) processed with our standard (Baldi et al., 2002) pipeline; *(b)* we replaced the counts, rates, fluxes and HRs for sources observed in adjacent (overlapping) fields (so called *duplicated*) with the *stacked* values according to the prescription of Chiappetti et al. (2006b), hereafter Report II; *(c)* we used more optical tables for identification as they became available in the meanwhile; and *(d)* fully reprocessed the identification in a more automated and objective way.

This report gives a short account of such procedure. For more details one can consult the web page `http://sax.iasf-milano.inaf.it/~lucio/LSS/ReIdent/procedure.html` which also contains a reference to similar pages for the previous 2 versions.

In section 2 I list the input database tables used as starting point for the identification, namely new X-ray data (2.1) and new optical-IR data (2.2), while in section 2.3 I give some information on the astrometric correction. The procedure is described step-by-step in the various subsections of section 3. In particular the ranking (see 3.3) is now generated almost automatically from positional probabilities (see 3.2), although after data verification and eventual visual inspection. Other sections give some simple statistics and the coverage of the different surveys (see 3.4) and a brief comparison with the XMM-LSS catalogue version 1 published in Pierre et al. (2007) (see 3.5).

We intend to use the updated complete catalogue as the source from which to extract subsamples for further work (e.g. Trinchieri, Giorgetti et al. in preparation), possibly up to a full photometric redshift catalogue.

2. Data sources

Our starting point has been the **same** *glorified correlation table* (GCT; table of pointers into all possible combinations of database tables, each one correlated with the `xmdsepic` table with a "standard" correlation radius or criterion) used in Report I, of which we made a copy.

This table has been then modified to include new X-ray sources or amended X-ray results (see 2.1), and references to tables in other (optical, IR) bands which became available after the end of Report I (see 2.2).

Table	Update	Content	History	(5)	(6)
`xmdsepic`	**Jan 07**	reference table with XMDS X-ray sources resulting from the Milan pipeline	used since Paper I and Report I; **new** added G12bis and stacking	n/a	
`xmdsdup`	n/a	clone of `xmdsepic`	n/a	$6''$	a
`nov06`	Nov 06	X-ray sources from the Saclay pipeline, band merged within $6''$	Paper I and Report I used `may05`; **new** this consistent with XMM-LSS catalogue (Pierre et al., 2007)	$10''$	b
`virphot`	Feb 06	VVDS "good" UBVRI(JK) photometry	used since Paper I and Report I	$6''$	c,z
`bad`	Jan 06	VVDS photometry flagged "bad"	as above	$6''$	c,z
`loiano`	Jan 06	VVDS Loiano U filter photometry	as above	$6''$	c,z
`vimos`	Jan 06	VVDS spectroscopic information	as above	$6''$	c,z
`sacphot`	Oct 03	CFH12K observations made by Saclay	used since Paper I and Report I	$6''$	d
`virradio`	Aug 03	entire VIRMOS1.4GHz catalogue	used since Paper I and Report I	$40''$	e
`radio`	Nov 03	entire XMM-LSS own ("Leiden") VLA radio catalogue	used since Paper I and Report I	$40''$	f
`specfup`	Sep 03	spectroscopy campaign of Oct 2002	used since Report I	$6''$	
`xlssc`	Jul 05	list of XLSSC clusters (Andreon's web site)	used since Report I	$2'$	g
`loto`	Jun 05	Lotoweb (Lyon LS3DB)	present since PaperI and Report I	$10''$	
`d1`	Sep 05	CFHTLS D1 field $u*g'r'i'z$ photometry supplied via IPAC in Feb 05	used since Report I	$6''$	h
`d1t3`	**Nov 06**	CFHTLS D1 field release T003 supplied by Saclay	**new!!**	$6''$	z
`w1`	Sep 05	CFHTLS W1 fields $u*g'r'i'z$ photometry supplied via IPAC in Jul 05	used since Report I	$6''$	h,z
`w1t3`	**Jul 06**	CFHTLS W1 fields release T003 supplied by Saclay	**new!!**	$6''$	z
`swire`	Sep 05	SWIRE data from above IPAC releases	used since Report I	$6''$	i,z
`swires05`	**Dec 06**	SWIRE Spring 05 public release (IRSA Gator)	**new!!**	$6''$	z
`ukidss`	**Feb 07**	UKIDSS DR1plus public release	**new!!**	$6''$	z
`galex`	Jun 04	unofficial FITS file with GALEX data	used since Report I	$6''$	j,z
`simbad`	Jul 05	SIMBAD sources	used since Paper I and Report I	$20'$	k,z
`ned`	Jul 05	NED sources	used since Paper I and Report I	$20'$	k,z
`usno`	Jun 05	USNO A2 catalog as kept at ST-ECF.	used since Report I	$6''$	j,z
`tajer07`	**Apr 07**	Tables A.1, B.1, B.2 from Tajer et al. paper	**new!!**	n/a	l
`polletta07`	**Apr 07**	Table I from Polletta et al. paper	**new!!**	n/a	m
`garcet07`	n/a	reserved for tables in Garcet et al. paper	planned		

Table 1. Database tables used as input to the present XMDS catalogue

(5) column (5) is the correlation radius used to populate the GCT with the object around the X-ray sources
(6) column (6) refers to the notes indicated below

a `xmdsdup` used to tag sources detected in overlapping fields (see Report I)
b correlation within $10''$ in uncorrected coordinates ; `nov06` and `may05` descend from same (May 05) Saclay pipeline run, only band merging is different
c all VVDS "authorized subsets" share a system of common identifiers, in particular `bad` and `virphot` are disjoinct sets, while the other two are subsets of the union of the latter.
d CFH12K observations made by Saclay outside of the VVDS area (or older data inside the VVDS area, identifiers shared with `virphot`)
e Bondi et al. (2003)
f Cohen et al. (2003)
g reverse correlation table
h see detailed notes in Report I. `d1` refers to the surroundings of Saclay Nov04 sources, `w1` to those of Saclay May05 ($ML > 20$) plus XMDS.
i see detailed notes in Report I. Most data come from the July 2005 IPAC release (flagged as `dataset=3`).
j table does not contain any scientific information, but is used only to assess whether it is worth asking to proceed for a more formal collaboration.
k SIMBAD and NED may also include data from some of our catalogues (radio and XLSSC).
l Tajer et al. (2007)
m Polletta et al. (2007)
z the update date indicated neglects the addition of G12bis counterparts in Jan 07.

2.1. New X-ray data

XMDS field G12 was not included in the original XMDS catalogue (`xmdsepic` table) because of the high background which made analysis impossible with our standard pipeline (Baldi et al., 2002). The field was reobserved during AO-5 (July 2006) as part of the latest XMM-LSS GO proposal (with ESA obs id 0404960501), and is the only one of the new fields to have been reprocessed with the Milan pipeline (across end 2006 and early 2007). The field is numbered 1112 in `xmdsepic`, and is rather shallow, containing only 36 new detections.

However these 36 detections required a lot of work, since one had to repeat (for them alone) the cross-correlation with all other existing tables, and the relevant cleanup, i.e. the *full* procedure described in Report I and Section 3. One had then to locate if they overlapped a detection in a neighbouring field (15 did), which detection to consider as primary (in 5 cases the G12bis one, in 10 cases the previous one), and to apply to all duplicated n-uples the stacking procedure described in Report II. Only at the end the results were appended to the GCT defining our catalogue.

After this the `xmdsepic` table comprises detections for original XMDS fields (G01 to G19, i.e. 1001 to 1019, excluding G12 and with G16a/b coded as 1116 and 1216), for the new field G12bis (1112), and stacked measurements (coded with `field=1000`).

We emphasize again that the stacking procedure (see Report II for details) does *not* imply a new (deeper) detection, hence it uses the positional information of the primary detection (chosen on the basis of best S/N, rate, flux, count and detection probability in the best band, as detailed in Section 3.3 of Report I). It is only the intensity related parameters (count, rates, fluxes, hardness ratio and associated errors and quantities) which are replaced (as detailed in Report II).

The change at level of GCT for the 173 sources detected in more than one field was to *repoint* the counterparts of the primary detection to the stacked measurements in `xmdsepic`. Hidden pointers with "technical autoranks" (8 and 5) preserve *a futura memoria* the association between the stacked source and its primary and secondary detections.

2.2. Input database tables

Table 1 gives a synoptic view of the database tables pointed from the GCT used for the present working catalogue.

Most tables where already present in the GCT used in Report I, and for them the counterpart association was not repeated (but the identification and ranking was repeated for all sources), except for what field G12bis is concerned (see 2.1 above).

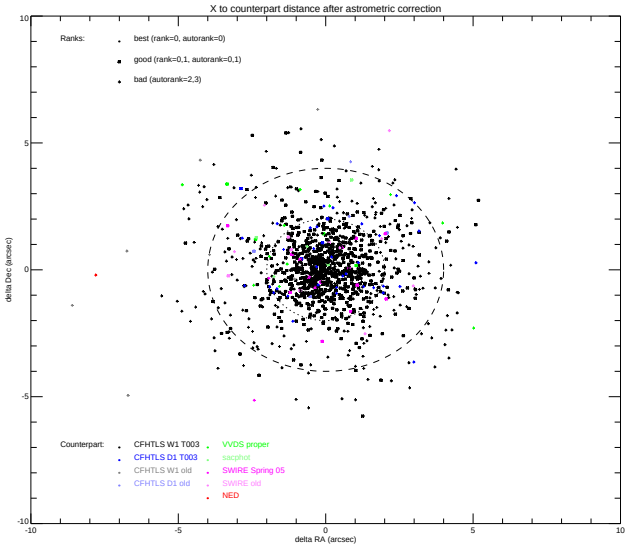

Fig. 1. Distances in RA and Dec between the X-ray corrected position and the best counterpart position. Different symbols indicate the identification quality. Only sources with rank=0 or 1 are plotted. A circle is plotted when the autorank is 0 or 1, and it is filled when both rank and autorank are 0, i.e. for the best candidates). A cross is plotted for lower autoranks (2 and 3) irrespective of rank. Different colours (as shown on figure) indicate the origin of the counterpart position for the distance calculation. Two fiducial radii of 2 and 4″ are also shown.

The tables which have been added ex-novo are those concerning CFHTLS release T003, the SWIRE Spring 05 public release, and the UKIDSS DR1plus public release, plus the tables referring to published papers.

For CFHTLS release T003, the files supplied by Saclay are more complete (see Section 3.2) in terms of area covered and magnitude limits. They have been ingested in temporary tables, and only the objects within 9″ from an X-ray source are kept online (the correlation was done however within 6″).

For SWIRE the public release Spring 05 is on one hand consistent with the data supplied by IPAC in July 05 (uses the same identifiers), on another it is not immediately comparable because for `swires05` I have chosen to use the recommended "aperture 2" fluxes (while `swire` uses now discouraged "aperture 3"). Also Spring 05 is more complete in terms of area covered and limit fluxes, but is more selective (includes only sources detected above a significance threshold, and at least at 3.6 and 4.5 μm). Therefore there are sources detected in single (or some) bands in `swire` which are missing in `swires05`. Data have been extracted at IRSA Gator within a radius of 10″ from a supplied list of positions, and then ingested. Data have been extracted from the band merged table (from 3.6 to 24 μm), analogous to `swire`, as well as from the 70 and 160 μm tables, ignoring the IRSA 24 μm standalone table. Correlation between 70 and 160 μm and the other bands has been done in Milan.

For UKIDSS the latest release available when the ingestion was done was DR1plus. Data have been extracted at WSA within a radius of 10″ from a supplied list of positions, and then ingested. They have been taken from the DXS and UDS surveys (the latter is centered on the Subaru fields and is absolutely marginal for us), and presently cover a nice rectangular strip across the XMDS (see Fig. 11). Most sources have a measurement in a single band (J or K), some in both. Later during 2007 WSA made available a further release (DR2plus). However since I had already done the correlations, the new release had altogether different id's while it added only two smaller, non-contiguous sky areas (one outside of XMDS), and almost no new information in terms of JK band coverage, I decided not to use it, and wait for a more complete future release.

The `tajer07` and `polletta07` tables are just an useful way to keep track of the results of the analysis in such papers (the Garcet et al. submitted paper could be handled similarly once accepted). They have been ingested from the LaTeX source at astro-ph. Since the data were derived from the "Report I" XMDS catalogue, the correlation on identifiers has been straightforward (for consistency with the new choices, sources with duplicated detections have been however repointed to the stacked entry).

Concerning results of the Saclay analysis, the pointers provided are to the full band merged physical table `nov06`, inclusive of spurious sources, and fields not considered for the `XLSS` catalogue in Pierre et al. (2007). We however provide a flag `xlsscat` telling whether the `nov06` entry is or not in the `XLSS` catalogue.

2.3. Astrometry

Readers are reminded that in Paper I we used the best (VVDS) identifications as input to the SAS task `eposcorr` to generate an astrometric correction (rigid shift) to be applied to all X-ray source positions (in the `xmdsepic` table, for the G fields covered by the VVDS.

In Report I the astrometric correction was recomputed for all G fields using the *pro tempore* best candidates (using in priority order CFHTLS `d1`, `w1` or VVDS positions). However the new correction, being consistent with the old one within the errors, was applied only to fields G09 and G14 to G19. The remaining fields still use the Paper I astrometry.

The only update regarding astrometry concerns field G12bis, which has been processed similarly to what described in Report I, but using in priority order CFHTLS `d1t3` (for most sources), `d1` or `w1t3` (for less than a handful).

We do not reproduce here a figure similar to Fig.4 of Report I, since the addition of the single G12bis point won't give much additional information.

I have instead produced a figure (Fig. 1) comparable with Fig. 9 of Paper I, which gives the distances in RA and Dec between the X-ray corrected position and the best counterpart position. By "best" I mean both that only entries with rank 0 or 1 are considered, but also that I selected as reference counterpart position the CFHTLS coordinates if available (in order W1 T003, D1 T003, old W1, old D1) then VVDS (i.e. `virphot` or `bad`), `sacphot`, SWIRE (first Spring 05 then the old table), `virradio`, `radio` and NED.

The results in term of positional accuracy are quite similar to those in Report I. In particular 97% of the sources have both RA and Dec offsets lower than 4″ , and 83% have both within 2″ (and no source has *both* offsets above 4″).

In terms of true distance 91% of the total is within 4″, which makes 97% of the good identifications (the circles in Fig. 1) and of the best (the filled circles in Fig. 1).

3. The procedure

As said above, the starting point for the present version of the catalogue is the *same* GCT generated by the procedure described in Report I for what concerns cleanup, pre-flagging and technical pre-ranking. The probabilities and final ranking was instead done afresh.

In practice this means that the counterpart associations for all tables, but the ones marked as **new** in Table 1 (i.e. CFHTLS T003, SWIRE Spring 05 and UKIDSS) has *not* been repeated, but simply *preserved from the previous version.*

The details of our procedure are given in `http://sax.iasf-milano.inaf.it/~lucio/LSS/ReIdent/procedure.html` or in internal notes.

Some of the steps listed there, and summarized below, have been in practice applied more than once at different stages (sometimes altogether repeated), not necessarily in the order in which they are listed.

3.1. Insertion of new tables

Unlike the brute force approach used in Report I (considering all possible combinations of counterparts given by the individual correlation tables with X-ray sources, and then doing a radical cleanup of spurious combinations), the addition of counterparts from the new tables (in order `d1t3`, `w1t3`, `swires05`, and `ukidss`) has been managed in an *incremental* way, exploiting the associations already made.

- a preliminary step is to create a pointer column for the new table in the GCT.
- then one *inserts a pointer* to the new table entry into existing counterpart sets when the object in the new table is closer to one of the existing counterparts in other table within a predefined radius. E.g. in the case of `d1t3` objects they were first compared with CFHTLS objects of the *earlier release of the same table* (here `d1`), then with the other tables of the same

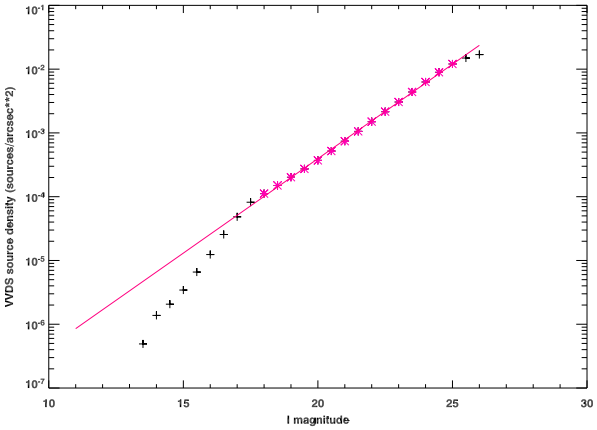

Fig. 2. Source count density for the VVDS. The range $18 < I < 25$ (in colour) has been used to produce the fits shown, whose parameters are given in Table 2.

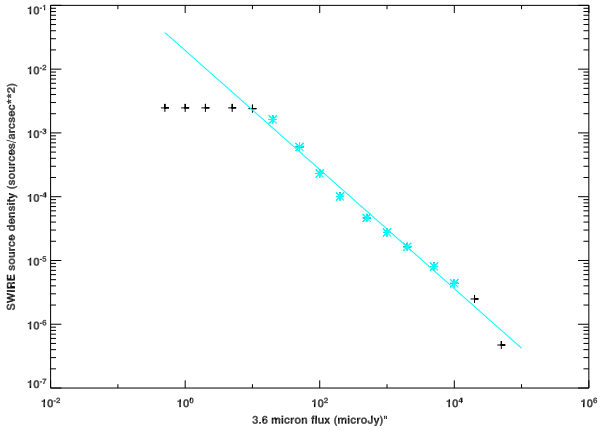

Fig. 4. $3.6\mu m$ source count density for SWIRE Spring 05. The (aperture 2) flux range indicated in colour has been used to produce the fit shown whose parameters are given in Table 2.

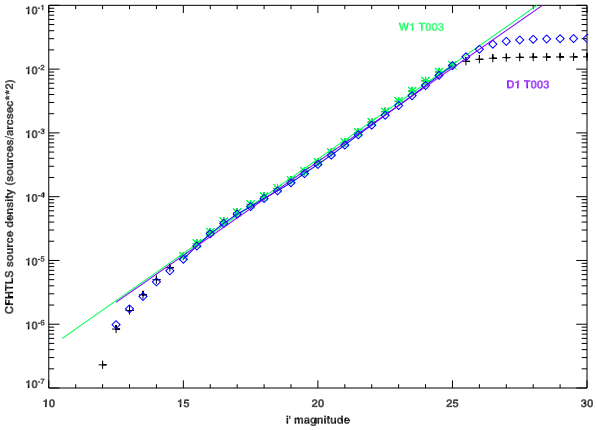

Fig. 3. Source count density for the CFHTLS D1 (crosses) and W1 (blue diamond) fields. The range $15 < i' < 25$ (in light colours) has been used to produce the two fits shown, whose parameters are given in Table 2.

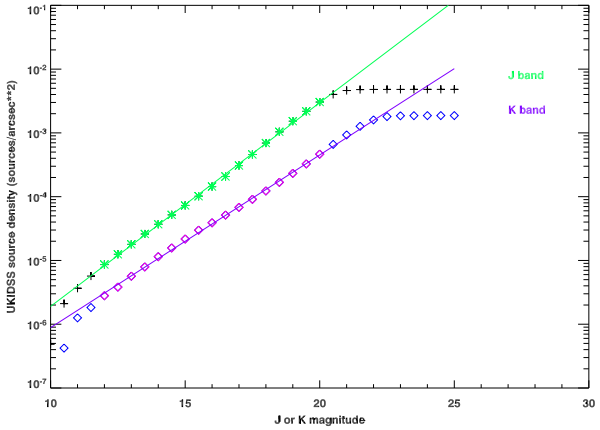

Fig. 5. Source count density for UKIDSS J band (crosses) and K band (diamonds) The magnitude range (12-20) indicated in lighter colour has been used to produce the fit shown whose parameters are given in Table 2.

origin (here for CFHTLS only old `w1`), then with other optical (VVDS, `sacphot`) or SWIRE positions, and finally radio or external catalogues.
- the correlation radius used has been $0.5''$ when comparing positions of the same origin (e.g. CFHTLS to CFHTLS), $1''$ when comparing to other optical or SWIRE catalogues, $1.5''$ in case of external catalogues, and taking into account the radio position error box for radio catalogues.
- then one *inserts a pointer* to the new table entry into previous GCT entries flagged as blank fields.
- finally *new counterpart sets* (new records) are inserted when an object in the new table is close to an X-ray source within the appropriate correlation table, but has no correspondence in the other photometric catalogues.
- next a verification of the pre-flagging is made for all modified entries (a case flagged "unique counterpart" can no longer be unique, or no longer a blank field, no longer the brightest, etc.). Also some of the flags and technical ranks can and have to be propagated to new entries.
- In particular in case of CFHTLS photometry one can reassign the star, galaxy or faint source and the saturated source flags. Since these are independent, a source can have multiple or contradictory flags (in particular can be saturated in VVDS but not in CFHTLS, and still be flagged saturated; or can be flagged as galaxy in one case and faint in another, considering that the magnitude threshold where the star/galaxy classification becomes undefined (the "faint limit") is I=21.5 for VVDS, i'=21.0 for CFHTLS).

Probability	m	density $n(brighter\ than\ m)$	a	b	tables
probvvds	I	$n(<I) = 10^{a+bI}$	-9.32636	0.29614	in order `virphot bad sacphot`
*probd*1	i'	$n(<i') = 10^{a+bi'}$	-9.32918	0.293859	for `d1t3` and `d1`
			-9.33087	0.296024	for `w1t3` and `w1`
probswire	F_λ	$n(>F_\lambda) = 10^{a+b*log(F_\lambda)}$			in order `swires05 swire`
	$\lambda = 3.6\mu m$		-1.70710	-0.933677	for `swires05` then `swire`
	$\lambda = 4.5\mu m$		-1.74473	-0.972905	then in order of λ for `swire`
	$\lambda = 5.8\mu m$		-2.06185	-0.911372	
	$\lambda = 8.0\mu m$		-1.66316	-1.04989	
	$\lambda = 24\mu m$		0.126906	-1.55492	
probukidss	J	$n(<J) = 10^{a+bJ}$	-8.76046	0.270712	taken best if both bands present
	K	$n(<K) = 10^{a+bK}$	-8.91420	0.319399	

Table 2. Parameters used for probability computation

The above procedures required a minimum of (sometimes manual) cleanup, in particular to deal with the tiling of adjacent CFHTLS W1 or SWIRE Spring 05 fields, or to deal with artifacts (spotted when the CFHTLS to VVDS distance was larger than $1''$).

We refer to Report I for the definition of *technical autoranks*, i.e. values of `autorank=-8,-3,6,7,8` mainly dealing with XMM field overlaps. In addition to the values defined there, `autorank=5` is used as a pointer to the primary detection of a stacked entry (see Report II). The entries with technical autoranks are not intended for scientific usage, so it is not worth giving further details.

3.2. Computing probabilities

Similarly to what was done in Report I, we computed the probability of chance coincidence between the X-ray source and its counterparts, based on the X-ray to optical (or IR) distance, the optical or IR intensity, and the density of sources brigther than a given intensity.

We computed `four` probabilities : *probvvds*, *probd*1 and *probswire* already present in Report I, and *probukidss* which is new. The latter however has been added later, and has not been used in the pre-ranking procedure.

All the probabilities have been *recomputed afresh*. They are based on a formula like

$$probability = 1 - exp(-\pi\ n(brighter\ than\ m)\ r^2)$$

where r is the largest between the X-ray to counterpart distance and $2''$ (i.e. the probability is *capped* to a distance corresponding to an intrinsic position uncertainty), and the density $n(brighter\ than\ m)$ is computed from simple linear fits as reported in Table 2. The same table indicates also the magnitudes or fluxes used to look up the density for the appropriate band.

The usage of *capped probabilities* (which essentially privilege intensity with respect to distance in choosing between multiple counterparts closer than $2''$ to the X-ray source) is new with respect to Paper I and Report I.

Probability *probvvds* is computed for sources with a VVDS counterpart (with preference to `virphot` then `bad`) and, but for the capping, is the same used in Paper I.

CFHTLS probability, traditionally called *probd*1, is computed for sources with a CFHTLS counterpart (in order `d1t3`, `w1t3` then `d1` or `w1`).

Probability *probswire* is computed in wavelength order. De facto all Spring 05 `swires05` sources have a flux at $3.6\mu m$ so the other wavelengths are used only for residual cases present at some bands in the old `swire` table.

Probability *probukidss*, in the case both (J and K) magnitudes are present (which is often not the case), is the best (smallest) of the two.

A probability of 99 ("undefined") is assigned whenever it cannot be computed.

The density of VVDS sources is interpolated from the same data and with the same formula used for Paper I (see Fig. 2).

The density of CFHTLS sources has been derived separately from the *totality* of the sources in the D1 T003 and W1 T003 data (ingested in a temporary table), with a coarse fit to the data (see Fig. 3), and has been applied also to the earlier releases of either field.

The density of SWIRE sources has been derived in each waveband from the *totality* of sources in the Spring 05 catalogue (using IRSA Gator in count-only mode) using aperture 2 fluxes (but applied to old `swire` aperture 3 ones, *sic!*): see Fig. 4 for $3.6mu$m (other bands not shown).

The density of UKIDSS sources has been derived separately for J and K bands from the *totality* of DXS data, using WSA in count-only mode: see Fig. 5.

The computation of density is based on source counts, but requires the knowledge of a sky area, which I computed as in Report I, using a grid of cells 0.01×0.01 degrees and counting how many cells contain at least one object. I obtained for D1 an area of 1.02 deg^2, for W1 12.03 deg^2, for SWIRE 7.84 deg^2 and for UKIDSS 8.95 deg^2.

With respect to Figures 1-3 of Report I one can note that the coverage at faint fluxes derived from entire catalogues (in Figures 2 to 4) is now virtually complete.

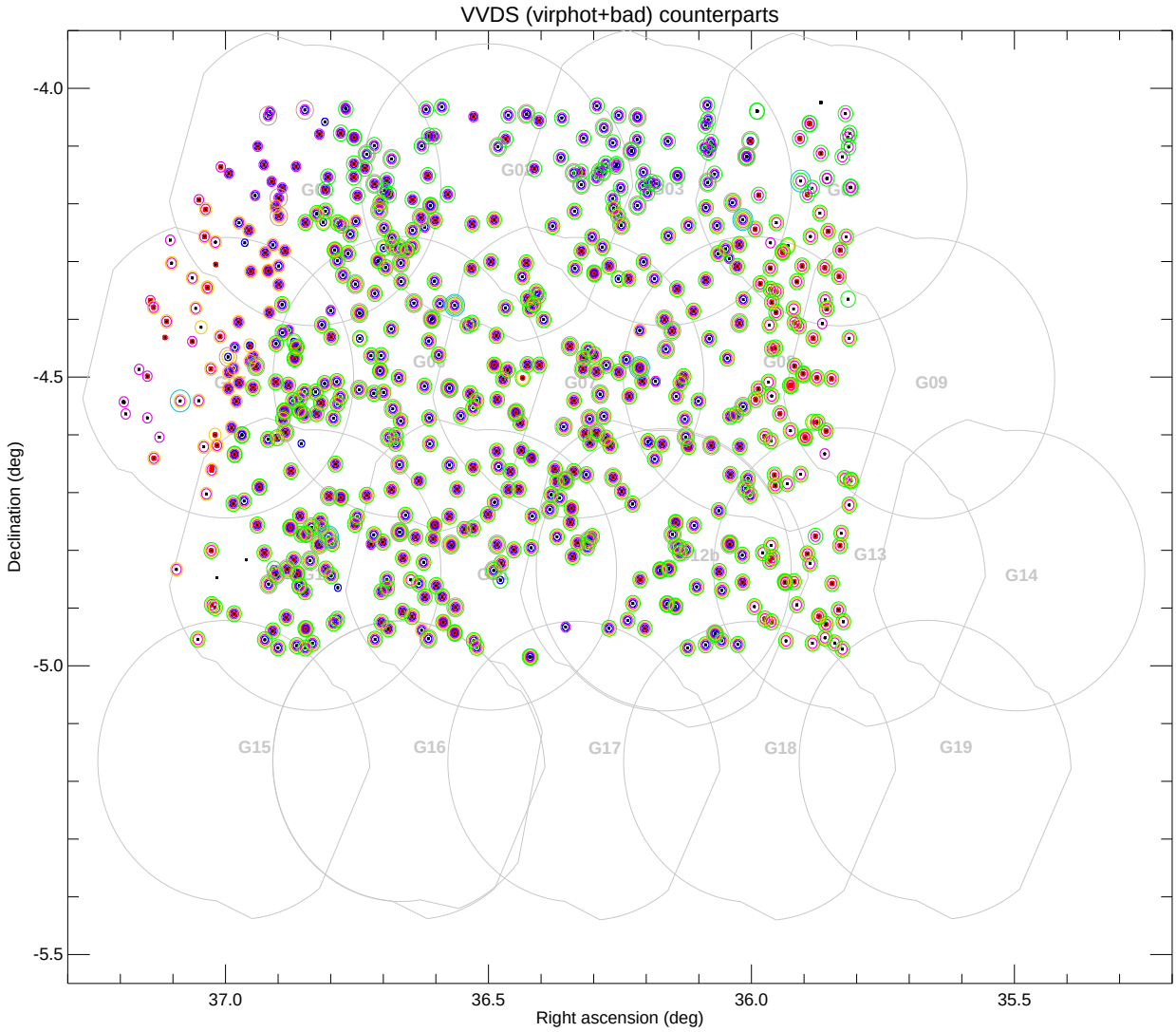

Fig. 6. Positions of the X-ray sources with a VVDS counterpart. For symbols see 3.4 in text. The VVDS covers almost entirely the top three rows of fields, with the main exception of the rightmost part of G04, G09, G13 and of G14. Note that the leftmost parts of G01 and G05 are covered only by the VVDS (and `sacphot` which here should be the same).

3.3. Ranking on probabilities

Once the probabilities have been computed, they can be used to assign *a priori* a preliminary rank (the *autorank*) to a particular association. For this I consider an individual probability p according to the following classification as in Report I :

- good if $p < 0.01$
- fair if $0.01 < p < 0.03$
- bad if $p > 0.03$
- undefined if $p = 99$

An *autorank=4* has been used to flag the blank fields (X-ray source is unidentified, has no counterpart).

For non-blank fields autoranks have been assigned based on the three "main" probabilities *probvvds*, *probd*1 and *probswire* as in Report I. An *autorank=0* has been assigned to the case where all three probabilities (or the largest number of probabilities which are not undefined) are all good. An *autorank=1* has been assigned when all non-undefined probabilities are at least fair (excepting of course those already ranked 0). Combinations are assigned *autorank=2* if not already ranked, and at least one of the probabilities is fair (but not all). Any entry where no probabilities are fair has received *autorank=3* if at least one probability is not undefined.

The remaining cases were also dealt with as in Report I i.e. I assigned *autorank=1* if the optical position is within the nominal X-ray error circle, *autorank=2* if it is within $4''$ and *autorank=3* otherwise.

A comparison of the new autoranks with those assigned in Report I was made, and things were found usually compatible, with the different choices due to the usage of capped probabilities. However autorank has to be con-

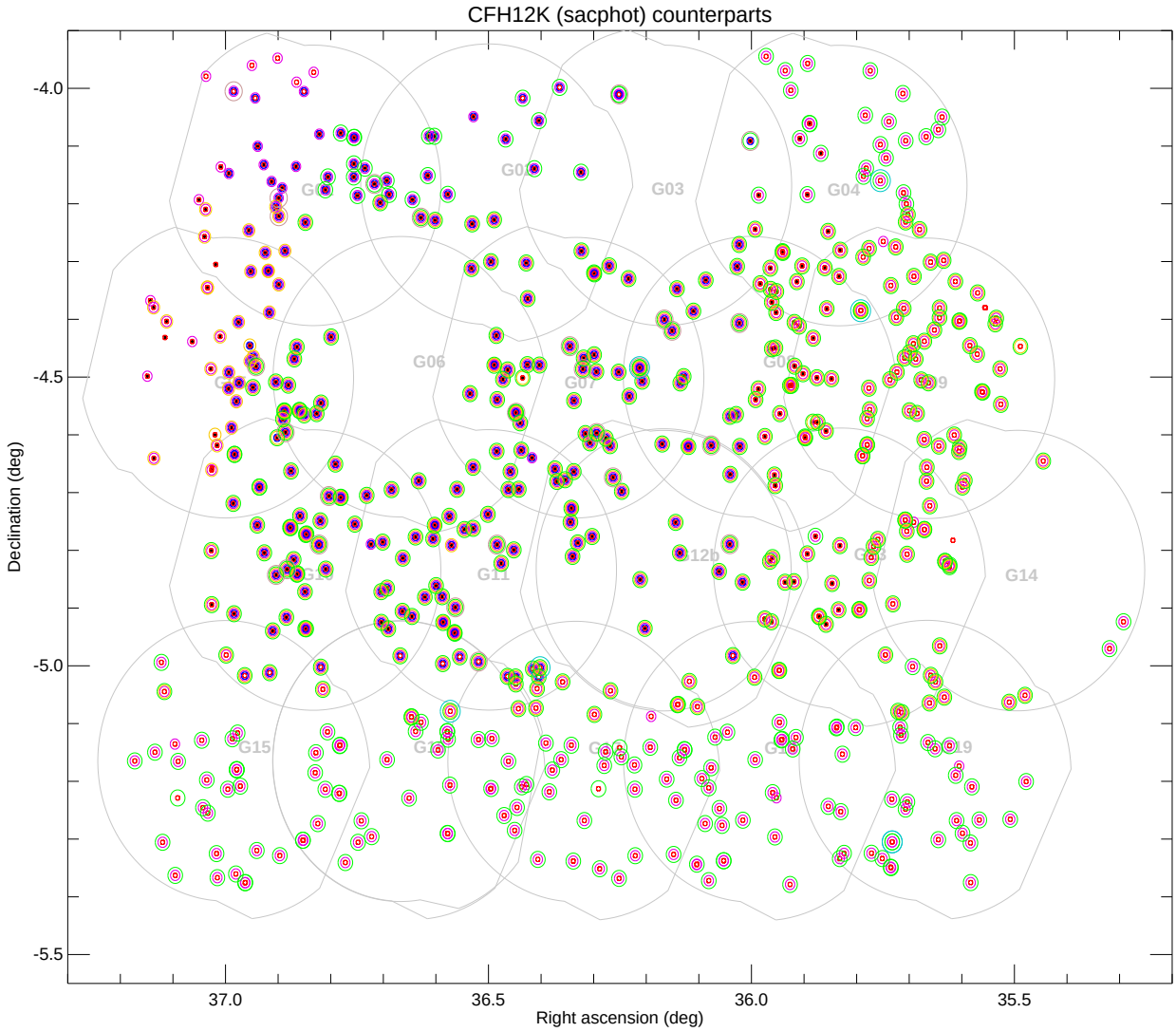

Fig. 7. Positions of the X-ray sources with a `sacphot` counterpart. For symbols see 3.4 in text. The reasons of the sparse coverage of the `sacphot` data, in particular the "holes" in G03 and G06 where VVDS data "coincident" with `sacphot` should exist, is unclear.

sidered only a working variable, and not used for serious data selection.

At this stage I passed from autoranks to ranks, defined according to the codes:

- rank -1 means *entry to be later totally removed*, i.e. technical autoranks, tiling overlaps, other duplicated or spurious cases
- rank 3 means *entry rejected* as counterpart
- rank 0, 1 and 2 indicate originally best, good or possible counterparts

The rank 0 is originally assigned to X-ray sources with a single counterpart with autoranks 0 or 1 (i.e. good or fair probabilities), and also to blank fields (autorank 4).

The rank 1 is originally assigned to the other X-ray sources with a single counterpart with worse autorank (probability),

In the case of X-ray sources with more than one possible counterparts, a tie-break on probability is made so that one preferred counterpart receives a rank 0 or 1, and all other secondary counterparts receive a rank 2. Therefore single rank 2 entries never occur alone. To allow easy spotting of X-ray sources with more than one counterpart, I systematically set flag 09 "ambiguous" for all entries corresponding to an X-ray source with more than one (rank 0-2) counterparts.

We than defined at an internal meeting the so-called *three Brera rules*, and associated *scores*:

- rule 1: at least one of the three main probabilities is good (score=1.0) or fair (score 0.5)
- rule 2: The X-ray source has a SWIRE `swires05` counterpart (score 1.0) or a `swire` one (score 0.5)

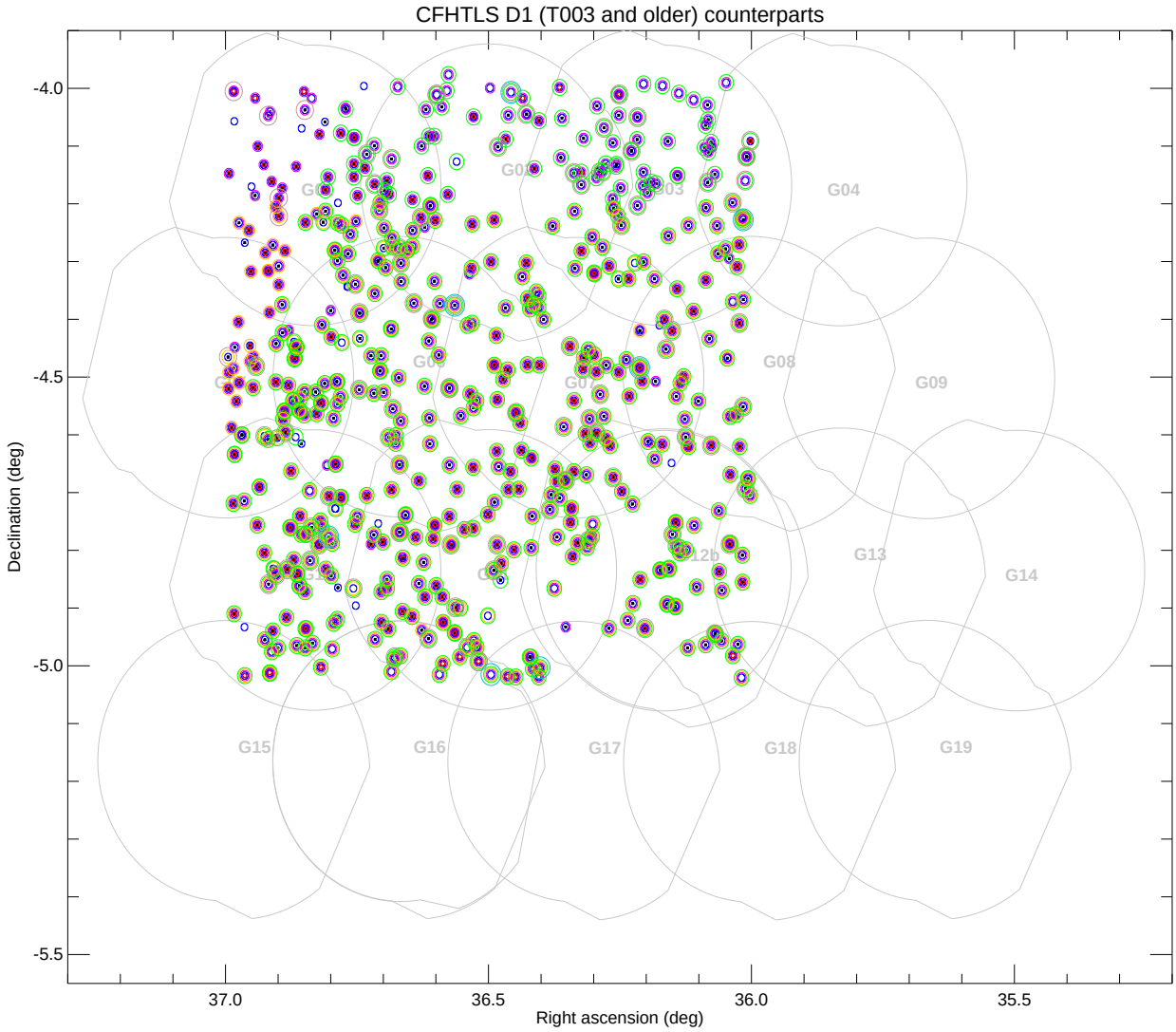

Fig. 8. Positions of the X-ray sources with a CFHTLS D1 counterpart. For symbols see 3.4 in text. The CFHTLS D1 covers the central part of the top three rows of fields. The former corner cut out in G01 (see Fig. 8 of Report I) was due to absence of SWIRE coverage in `d1` data got via IPAC. Now `d1t3` covers the entire square degree.

– rule 3: The counterpart set with the best (smallest) probability (considered as the smallest of the three main ones) has a probability ratio of at least 10 with respect to all other possible counterparts (for verification purposes this is done considering rank 2 and 3 counterparts). If rule is met, a score of 1.0 is added.

An inspection of the data has verified that the best counterpart has a total score (which can then range from 0.0 to 3.0) greater or equal to all other possible counterparts. In the few cases (just 6) where this did not occur, this was an indication of mistakes which were corrected (except for a marginal case where all counterparts are very weak).

The score was also used for a systematical tradeoff of *ambiguous* cases, i.e. those with more than one possible counterparts.

It has to be noted that the data inspection was done in all cases looking at the probabilities and the other data accessible within the GCT. In most cases, whenever necessary, this implied also a visual inspection including overlaying region files over VVDS or DSS-II thumbnail images as described in Section 3.7 of Report I.

Any peculiarity was always noted in the "comments" associated to the source entry.

Most of the above analysis was done before the UKIDSS catalogue was inserted, and therefore *probukidss* was not used for ranking and scoring. However it shall be noted that of the 782 X-ray sources falling within the UKIDSS area (see Fig. 11), for 709 (i.e. 91%) their best (rank 0-1) counterpart has been observed also by UKIDSS. So the sensitivity of the various surveys seems nicely matched. The UKIDSS counterparts have been con-

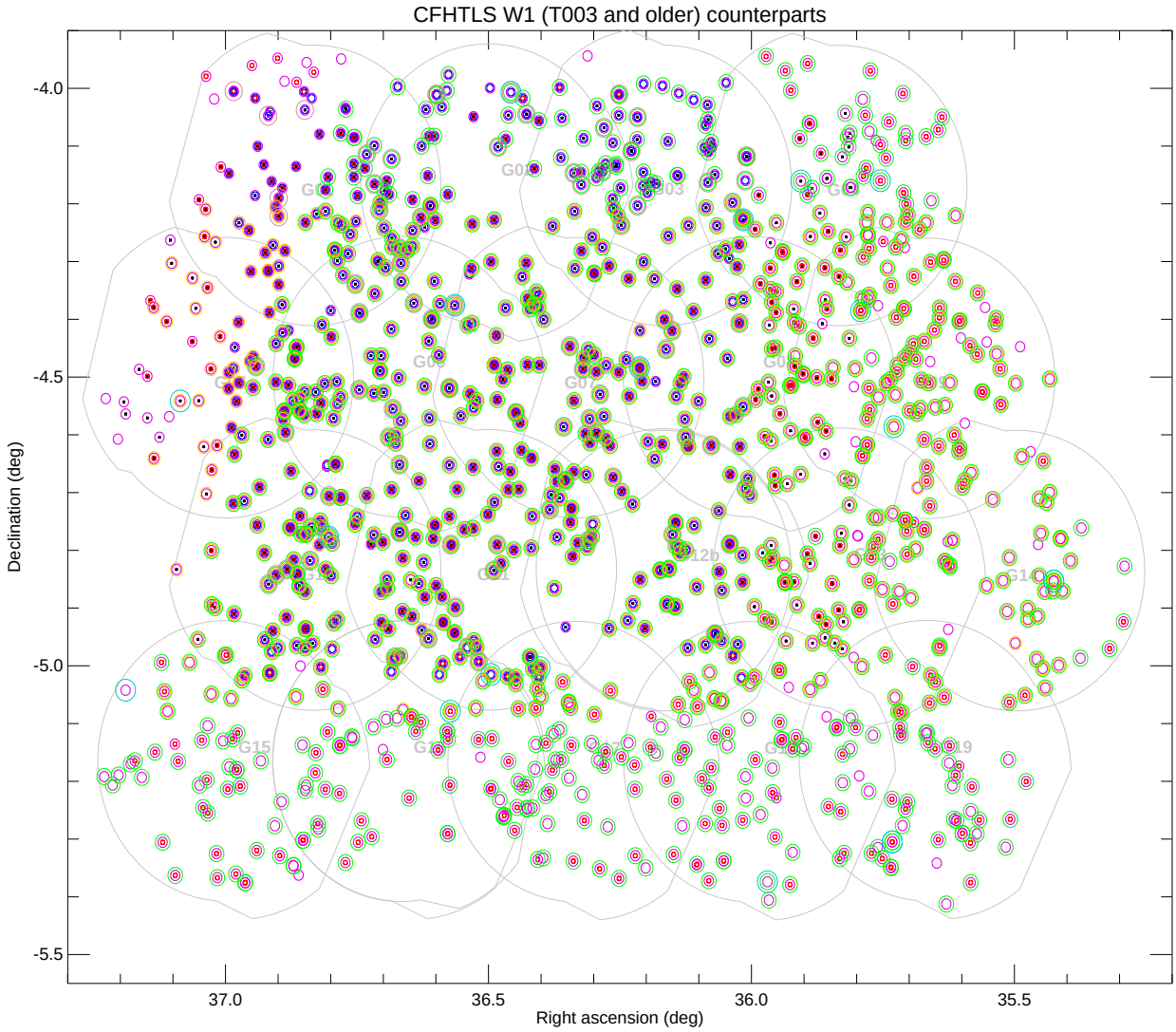

Fig. 9. Positions of the X-ray sources with a CFHTLS W1 counterpart. For symbols see 3.4 in text. The CFHTLS W1 covers all fields also outside the D1 area. Here too `w1t3` gives a complete coverage, while older `w1` was affected by absence of SWIRE coverage in data got via IPAC (see Fig. 9 of Report I).

sidered during data inspection, and in particular for the analysis of the ambiguities.

The result of the final inspection of ambiguous cases has been the following :

- A small number of cases (46) considered *subambiguous* has been resolved. They are classified with the other unambiguous sources, i.e. have a *single counterpart* with *rank 0 or 1* (for all unambiguous sources there is really no difference between the two ranks). The "ambiguous flag" 09 is **not** set. For these objects the flag 09 remains set for the rejected (rank -1) entries for future record. Comments mark peculiarities. For instance this category includes old `swire` 8 μm only sources, which are said to be contaminating asteroids.
- *Ambiguous cases with a definite preference* occur when the score of one counterpart is definitely better than all the other. In such case the best counterpart receives systematically `rank=0`, the other have `rank=2`, and flag 09 is set for all.
- *Really ambiguous* cases are those where the score does not allow to choose between the counterparts, In this case the best counterpart is selected purely on probability and receives systematically `rank=1`. As above the other have `rank=2` and and flag 09 is set for all.

I recall that in the database one can combine rank, (autorank) and flag to select counterparts by quality. E.g. an expression like `rank=0 and not find_in_set('09',flags)` locates the single rank 0 sources (the best), while `rank=1 and`

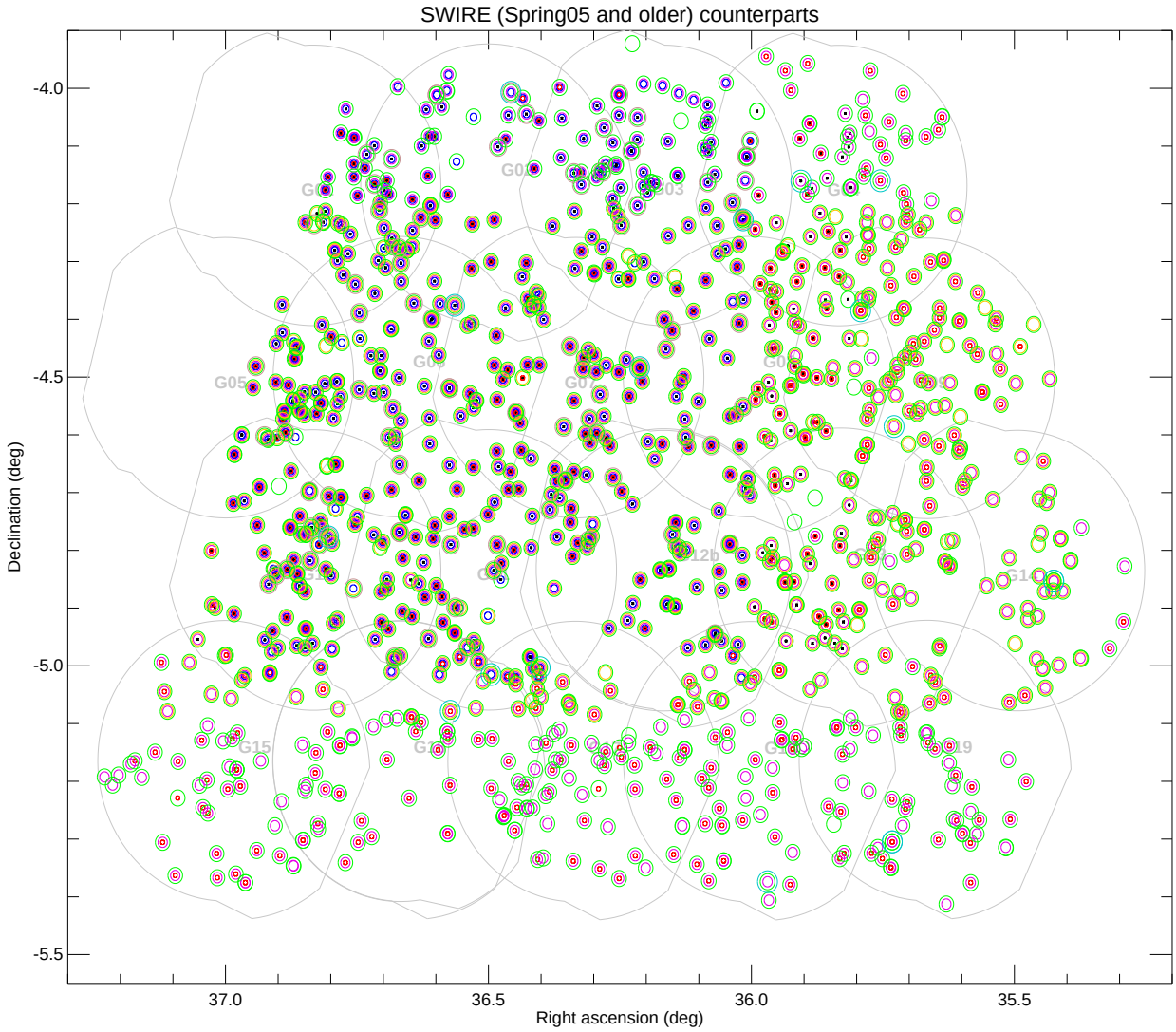

Fig. 10. Positions of the X-ray sources with a SWIRE counterpart. For symbols see 3.4 in text. SWIRE covers almost all the fields, but excludes most of G05 and G01.

`find_in_set('09',flags)` locates the best counterpart of the "really ambiguous" cases.

3.4. General properties

At the end of the identification procedure, the catalogue contains 1356 valid entries (i.e. those with rank 0-2) corresponding to 1168 distinct X-ray sources.

Namely there are 856 rank 0 and 312 rank 1 identifications (each one corresponding by construction to a single X-ray source). In addition there are 188 entries with rank 2 corresponding to 162 distinct X-ray sources (79 of them have a rank 0 entry, i.e. are "0/2 ambiguous" and 83 have a rank 1 entry, i.e. are "1/2 ambiguous"). The number of ambiguous sources improved (decreased) with respect to Report I.

6 of the rank 0 entries have `autorank=4`, i.e. are blank fields (X-ray sources with no counterpart). There is an entry currently flagged as rank 1 and autorank=4 which is not a blank field, but spoiled by a bright uncatalogued star (X-ray source #624). The number of blank fields improved (decreased) with respect to Report I also due to UKIDSS

I summarize some more statistical information in Table 3, giving a breakdown by significance, identification reason, autorank or kind of counterpart(s).

One can notice in the central part of the table the way the various optical surveys extend each other, which can be examined also in conjunction with the plots presented below in Fig. 6 to 12.

Of the 10 objects without an optical or SWIRE identification (the difference from 1158 in the fourth row from bottom of the first part of Table 3 to the total of 1168), 6 are the already discussed "pretended" blank fields, one is the contaminated #624, and 3 are in fields with UKIDSS counterparts only.

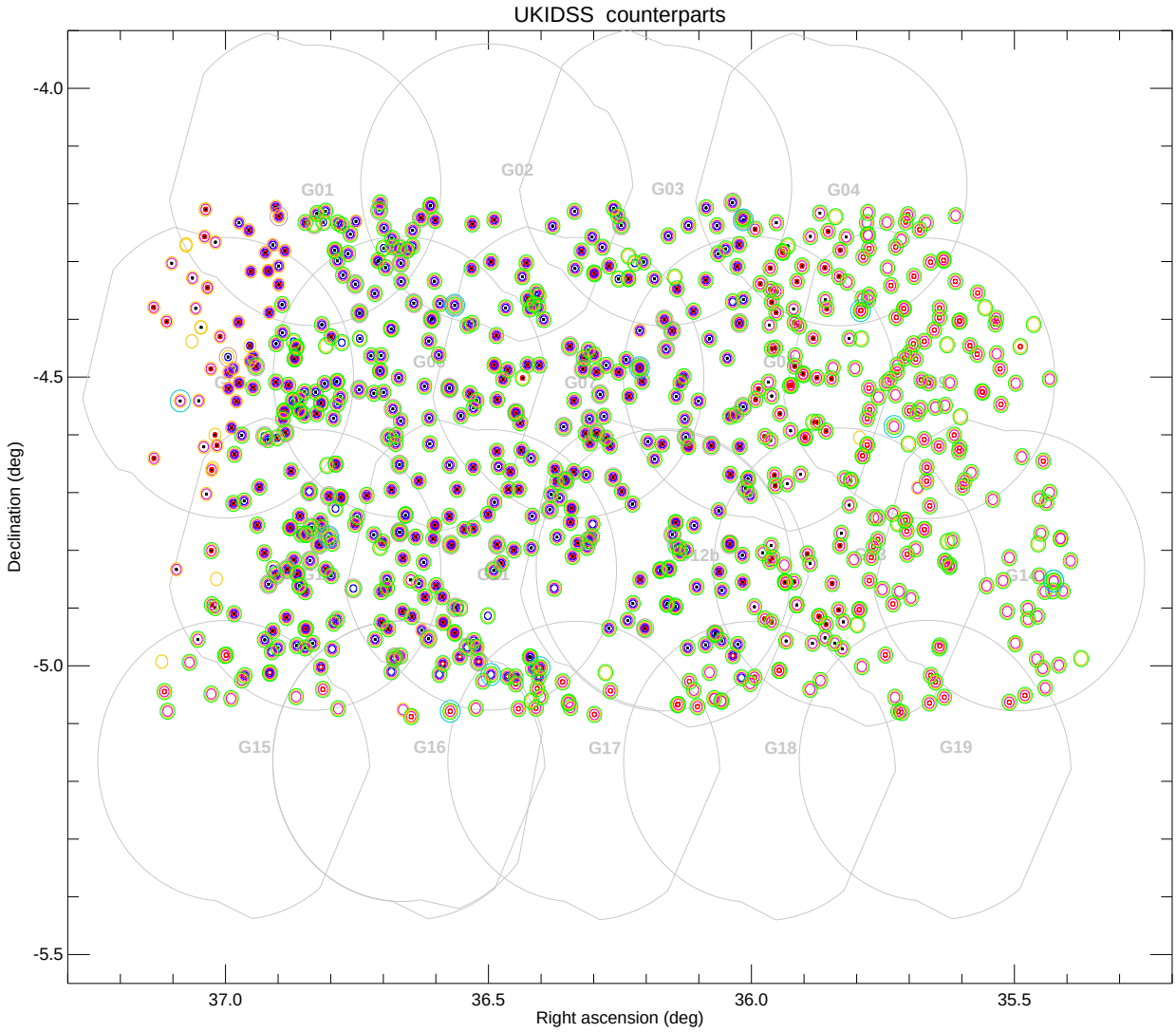

Fig. 11. Positions of the X-ray sources with a UKIDSS counterpart. For symbols see 3.4 in text. UKIDSS DR1plus covers only a strip along RA in the XMDS area. DR2plus (not shown and not considered) extends also to two non-contiguous areas (one in G03-04 and the other one north of it outside the XMDS).

I note also that, not obvious from the table, there are also 26 objects which have a SWIRE counterpart without an optical counterpart in our catalogues, which reduce to just 9 without even an UKIDSS counterpart. Some of them have flag 02 ("weak sources") set (but not when the association is with a relatively bright SWIRE source and absolutely unambiguous). All of them have *just* a SWIRE counterpart, in 3 cases this is only an old `swire` counterpart only, so they could be classified as "weak source" or "blank" fields (#289, #833, #1271), but the other are confirmed in Spring 05 (#135, #303, #840, #968, #1337) and #383 corresponds with an uncatalogue enhancement in the optical finding chart.

I note that there are 242 objects with a GALEX counterpart, concentrated in the three field rows G01/02/03, G05/06/07 and G10/11 (where more than 50% of sources has a GALEX counterpart). Since I do not have information about the full coverage of the GALEX observations, nor whether the GALEX list used is up to date, I'm not in a position to comment whether this is of interest to solicit further formal contacts.

The top part of the tables classifies the identifications by the various working criteria used, including the autoranks, scores and "Brera rules" quoted above.

One can also use a more conservative criterion in assessing the identification quality, similar to the one we plan to use for the future AGN samples. This is shown at the bottom of Table 3. One can consider only the latest or stable release of the main catalogues (T003 for CFHTLS and Spring 05 for SWIRE), while all other are considered merely ancillar. A wide majority (1130 out of 1168) of our sources has a counterpart in those catalogues, and of them 57% has a good probability, and 84% a good or fair one. So these 951 cases can be considered a conservative

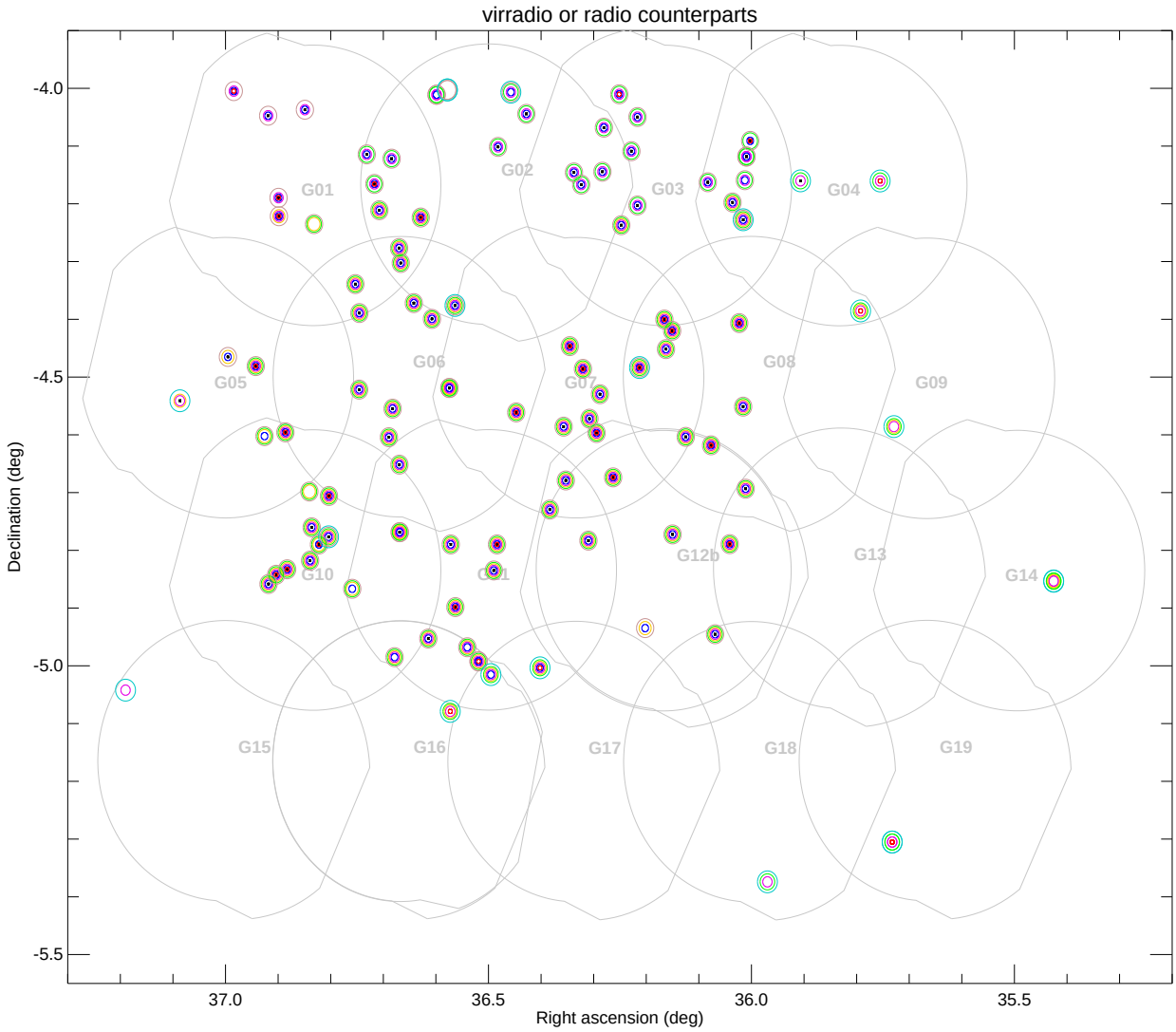

Fig. 12. Positions of the X-ray sources with a radio counterpart at 1.4GHz or at 74 and 325 Mhz. For symbols see 3.4 in text.

sample. Of the 179 sources which have a bad probability, 5 have a fair VVDS probability, and 16 (2) a fair (good) UKIDSS probability (so they won't change much). It shall be noted that the goodness of the identification is unrelated with the ambiguity. 97 of the 648, and 134 of the 951 are flagged as ambiguous, but only 20 of the 179.

Fig. 13 represents the distribution of the probabilities defined in Section 3.2. One can see that the choice of the good, bad and fair probability ranges made in 3.3 is sound.

It is very useful to evaluate whether in a given region we do not find counterparts in a given table because either they do not exist or the region has not been observed. I include 6 figures (from Fig. 6 to Fig. 12) which give the sky areas covered by the various surveys used by us overplotted on the footprint of the FoV of our fields. Each figure lists *only* the (autorank 0-1) sources with a counterpart in a *given* table (i.e. a non null entry in the GCT). The symbols used indicate in which other tables there is *also* a counterpart.

Such symbols are concentric circles of different colours, corresponding from the inner to the outer to :

- a small black dot indicates a VVDS counterpart (`virphot` or `bad`, thus any circle with the centre filled is also a VVDS source
- a small red circle indicates a `sacphot` (CFH12K) source
- a slightly larger blue circle indicates a CFHTLS D1 source (T003 or older)
- an even larger magenta circle a CFHTLS W1 source (T003 or older)
- a larger orange circle an UKIDSS DR1plus source
- a larger green circle a SWIRE source (Spring05 or older)
- a larger pink circle a `virradio` source
- the largest cyan circle a `radio` source

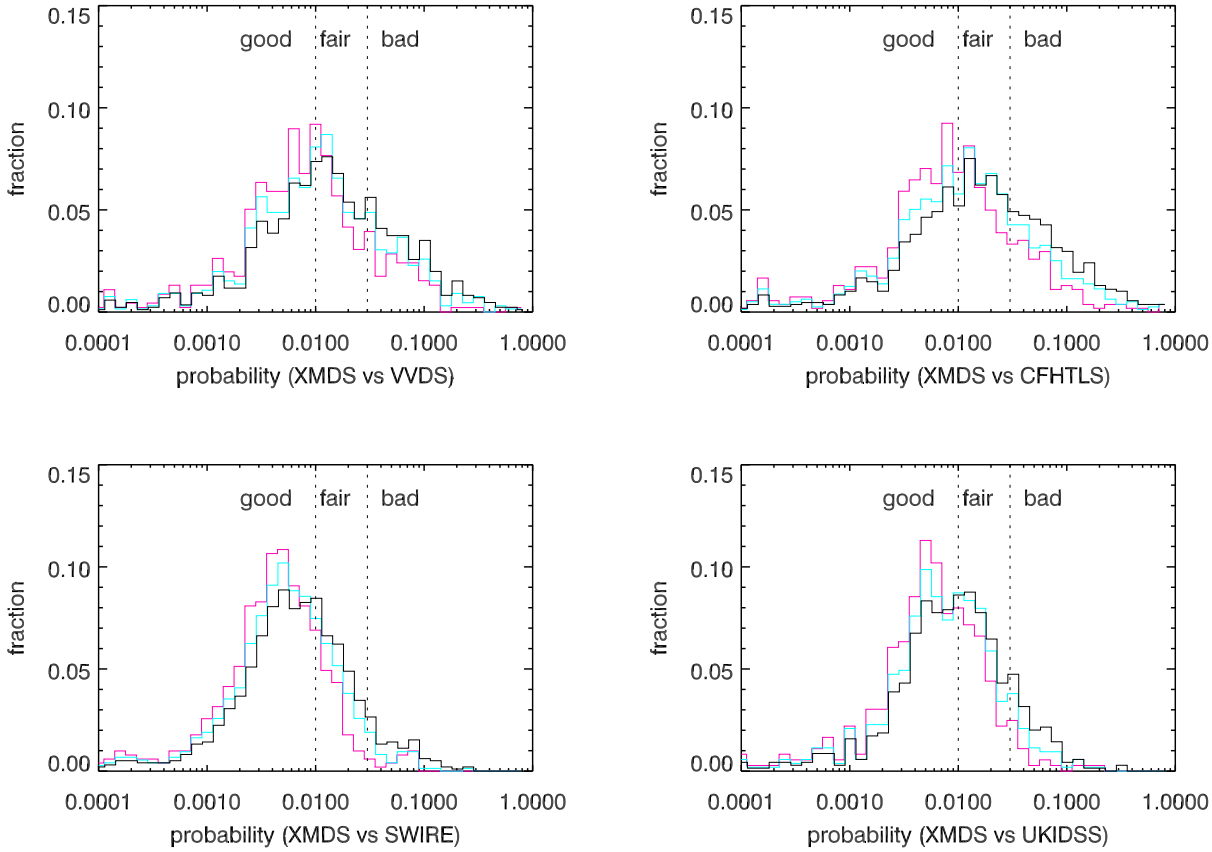

Fig. 13. Histograms of the four probabilities (*probvvds*, *probd*1, *probswire* and *probukidss* normalized to the total number of rank 0-1 objects with not undefined probability in the total sample (black), with a significance of at least 3σ in any band (cyan), or of at least 4σ (magenta). The dashed fiducial lines identify the loci with good, fair, or bad probability.

The issue of incomplete sky coverage for SWIRE-related `d1` and `w1` table raised in Report I seems now solved.

3.5. XMDS vs XMM-LSS

This section is partially based on the revision of some material prepared for, but not included in, the XMM-LSS catalogue paper (Pierre et al., 2007).

We remind here the main differences between the XMDS and the Xamin (Saclay) pipeline. *(a)* XMDS uses the SAS to detect candidates in 5 energy bands simultaneously operating on event files merged from all 3 XMM cameras and from the entire XMM field of view, and *(b)* then applies the Baldi et al. (2002) characterization. *(c)* The event pattern selection, *(d)* the removal of redundant sources and *(e)* the astrometric correction are also different.

I did a quick comparison between our XMDS sources and the sources in the `nov06` table derived from the Saclay pipeline, and in particular those included in the published XMM-LSS (`XLSS`) catalogue (Pierre et al., 2007). The latter can be selected by the mysql condition `xlsscat=1` applied to the GCT.

Of our 1168 XMDS sources, 1082 have a `nov06` counterpart, and in 943 cases they are in `XLSS`. Of the remaining 139, 56 are flagged as spurious on the basis of the Saclay detection likelihood (so are not eligible for `XLSS`). The other 83 are not spurious but some (17) fall in XMM fields considered not eligible for `XLSS` (mostly G12, while G12bis has not yet been analysed in `XLSS`, but also B04 and B32). The rest is due to different choices when resolving the overlap between adjacent fields.

Of 86 XMDS sources which are not in `nov06`, 49 are at high ($> 10'$) off-axis angles. Some of the other are ultrasoft (band A) sources (with such band not considered in the Saclay pipeline) or very weak sources anyhow.

As said above the XMDS catalogue includes 1168 (non redundant) X-ray sources above its probability threshold, while the present `XLSS` catalogue has 1574 X-ray sources in the G fields (but G12 is excluded from the official catalogue and G12bis has not yet been analysed). We correlate the two catalogues within a distance of $10''$.

Objects	Total	$> 4\sigma$	$> 3\sigma$	$> 2\sigma$		
detections[a]	1358					
independent sources	1168	576	851	1165		
Condition	Total	unique	B & C[b]	blank fields	other	
rank 0 unambiguous	777	56	553	6[c]	162	
rank 1 unambiguous	229	23	77	1[d]	128	
rank 0/2 ambiguous	79	0	42	0	37	
rank 1/2 ambiguous	83	0	20	0	63	
Condition	Total	autorank=0	autorank=1	autorank=2	autorank=3	
rank 0 unambiguous	777	319	305	147	0	
rank 1 unambiguous	229	1	11	53	164	
rank 0/2 ambiguous	79	40	32	6	1	
rank 1/2 ambiguous	83	20	21	23	19	
Condition	Total	score=3.0	2 to 2.5	1.5	1.0	0 to 0.5
All unambiguous	1006	390	266[e]	143	81	126
All ambiguous (best candidate)	162	19	92	22	9	20
Condition vs "Brera rule"	Total	1 good	1 fair	2 `swires05`	2 `swire`	3 prob.ratio
All unambiguous	1006	563	829	794	861	547
All ambiguous (best candidate)	162	99	137	130	140	28
Rank 0 and 1 identifications						
with VVDS counterpart	600		+			
with CFHTLS D1 counterpart	531	+				
with CFHTLS W1 counterpart	1075	=				
with either D1 or W1		1115	=			
with either VVDS or CFHTLS			1127			
with VVDS, CFHTLS or `sacphot`				1132	+	
with SWIRE Spring 05 counterpart	924					
with any SWIRE counterpart	1001				=	
with optical or SWIRE counterpart					1158	+
with UKIDSS	711					=
with UKIDSS only	3					
with optical, UKIDSS or SWIRE						1162
Rank 0 and 1 identifications dupdition	Total	$p < 0.01$	$p < 0.03$	$p > 0.03$		
with counterpart in		good	or fair			
W1 T003	1068					
W1 or D1 T003	1104					
W1 or D1 T003 *and* SWIRE Spring 05	898					
W1 or D1 T003 *or* SWIRE Spring 05	1130	648	951	179		

Table 3. Basic statistics ot the present XMDS catalogue

a at $p < 2 \times 10^{-4}$
b brightest and closest
c blank fields flagged as autorank=4 or flag 01 set, see text
d affected by bright uncatalogued star (#624, see text)
e including 96 solitary counterparts for which max score is 2

There is a number of XMDS detections (in the raw `xmdsepic` table), not appearing in `XLSS`, of which, if one correctly neglects stacked entries (since `XLSS` processes individual fields alone), 63 have an off-axis angle larger than $13'$ and so are excluded by construction by `XLSS`. The same occurs for another group of 26 XMDS sources which are very soft (detected with some significance only in band A, 0.3-0.5 keV, not used by `XLSS`). Most of the other 48 are quite faint and at large off-axis angles, but for some of them it could be worth investigating.

There is a larger number (557) of `XLSS` sources not appearing in XMDS. They include 30 extended sources (5 C1 and 14 C2 clusters plus other 11 detected as extended only in the hard band) for which XMDS is not optimized, while all the others are sources at low significance, consistently with the highest threshold of XMDS (see the cross-calibration on common sources below). Namely 149 sources with the highest likelihood in band B, and 65 in band CD have a detection likelihood below 20. Only 60 have their best likelihood above 40, and only 18 above 75.

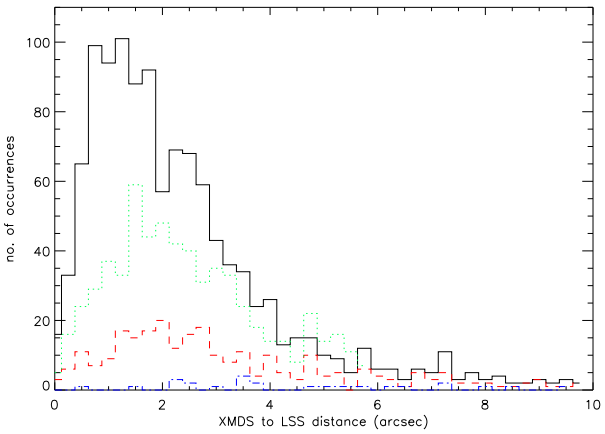

Fig. 14. Histogram of the angular distance between XMDS and XLSS X-ray positions in astrometrically corrected coordinates. The solid (black) line refers to all sources. The long dash (red) line to the source having a detection likelihood below 40 in the best band in which they are detected. The dash-dot (blue) line to extended sources. For comparison we report also the distribution of the inter-band distance `maxdist` between XLSS positions in bands B and CD, for sources detected in both bands (dotted line, green).

The number of common detections between the two catalogues is nominally 966 detected in the same field, but this value can be raised to 1149 taking into account the different handling of redundant detections (which each catalogue can, according to its own criteria, remove in different pointings). We concentrate now on the comparison of such common objects.

The distance between the (astrometrically corrected) X-ray positions in the two catalogues is shown in Fig. 14. 51% of the sources are closer than $2''$, 85% closer than $4''$ and only 6% are more distant than $6''$, in general concentrated among the sources with lesser significance, and the few extended ones. The agreement between the XMDS and `XLSS` positions, peaking around $1''$, is better than the typical inter-band distance between `XLSS` detections in the two energy bands, which peaks around $2''$.

Fig. 15 attempts to cross calibrate the detection likelihood of `Xamin` with the chance probability of the XMDS (for definition see Baldi et al. (2002)).

Alternatively one can use Fig. 16 to cross calibrate the detection likelihood of `Xamin` with the significance in terms of number of σ of the XMDS (see Paper I and references therein). One can see that a likelihood of 75 corresponds more or less to the 4σ level, and one of 40 to the 3σ level.

The common subset of 1149 sources includes 26 potentially extended sources (15 classified C1, of which 5 detected in both bands and 1 detected only in the hard band but satisfying the extension (C1) criterion in such band); 11 classified C2 of which 2 detected in both bands, and 4 detected only in the hard band but satisfying the C2 criterion in such band).

We also note that 88% of the common sources have B as the best band (highest likelihood) in `XLSS`. 93% of the sources are observed by `XLSS` in the B band, 62% are observed in the CD band, and 55% are observed in both. This can be compared with the totality of the `XLSS` catalogue where 85% of the sources have B as the best band, 89% are observed in the B band, 48% in the CD band and 37% in both (there is no appreciable difference between the full catalogue and the sources in the G fields alone). The XMDS by construction includes measurements in all 5 energy bands even if the source is above the probability threshold only in one. If we consider good detections for XMDS only those with $p < 2 \times 10^{-4}$ in the band, we have that 92% of the common sources are detected in the B band, 39% in the CD band, and 34% in both.

To compare the count rates in the two catalogues, one has to note that XMDS operates on camera-merged event files, and therefore computes automatically a MOS1+MOS2+pn rate, while `Xamin` works on MOS1+MOS2 and pn separately. Therefore we compare the XMDS rate with an `XLSS` camera merged rate computed as

$$rate = \frac{rate_{MOS}(exp_{MOS1} + exp_{MOS2}) + rate_{pn} exp_{pn}}{exp_{MOS1} + exp_{MOS2} + exp_{pn}}$$

(**This rate is not the same as the plain summed rate used elsewhere in Pierre et al. (2007) !**)

The count rates are compared in Fig. 17. They match reasonably well, although their average ratio is not unity (which is not surprising considering they result from independent processing, and in particular the event pattern selection used in XMDS and `XLSS` is different), but the `XLSS` rate is 1.092 times higher than the XMDS one in the B band, and 1.192 times higher in the CD band.

It can be seen that most outliers are either concentrated at low rates (i.e. poor significance) or correspond to extended sources, for which the XMDS obviously fails in characterizing the source.

The fluxes, computed for XMDS according to the prescriptions of Baldi et al. (2002) and for `XLSS` as explained in Pierre et al. (2007), are compared in Fig. 18. Extended sources classified C1 are excluded as their flux is set to undefined. The fluxes match qualitatively, although there is a systematic difference : namely the `XLSS` fluxes are 0.848 lower than the XMDS fluxes in the B band, while they are only 0.95 lower in the CD band. This despite the different agreement between rates quoted above.

It can be seen that a larger scatter in fluxes occurs for the sources which have poorer significance in either catalogue, while outliers are generally due to sources presumably falling near a chip gap on one detector (and as such characterized by a `fluxflag` of 2), or exceptionally by residual C2 extended sources (for which the `XLSS` flux is computed from the pointlike rate).

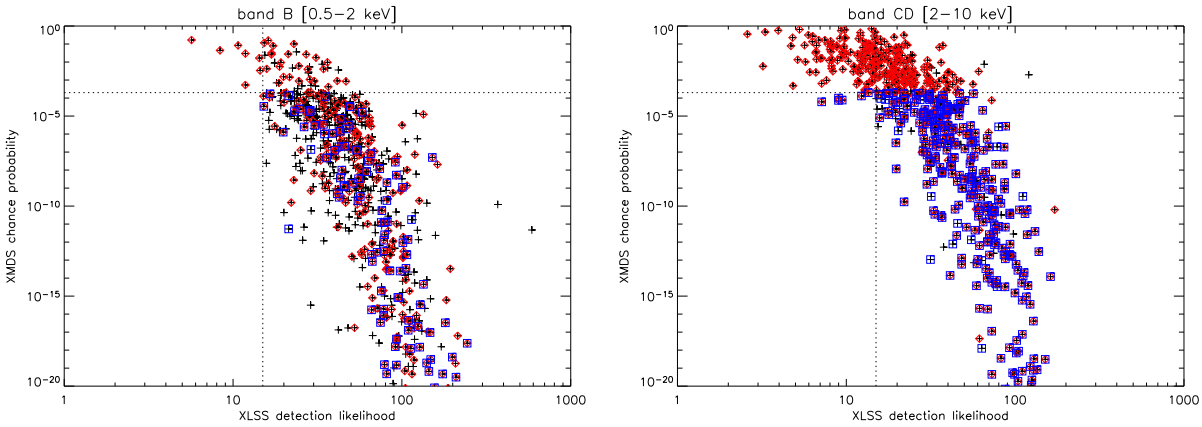

Fig. 15. Cross calibration between the `XLSS` (Xamin) detection likelihood and the XMDS chance detection probability. Left panel for the soft band, right panel for the hard band. The dashed lines indicate the two acceptance thresholds of $ML > 15$ and $p < 2 \times 10^{-4}$ (remember that a source can be accepted if it is above the threshold *in at least one band* but not necessarily in all). Crosses indicate all objects detected in the given band. A (red) diamond surrounds the sources detected above threshold *in both bands* in the `XLSS`, while a (blue) square surrounds those detected above threshold *in both bands* in the XMDS.

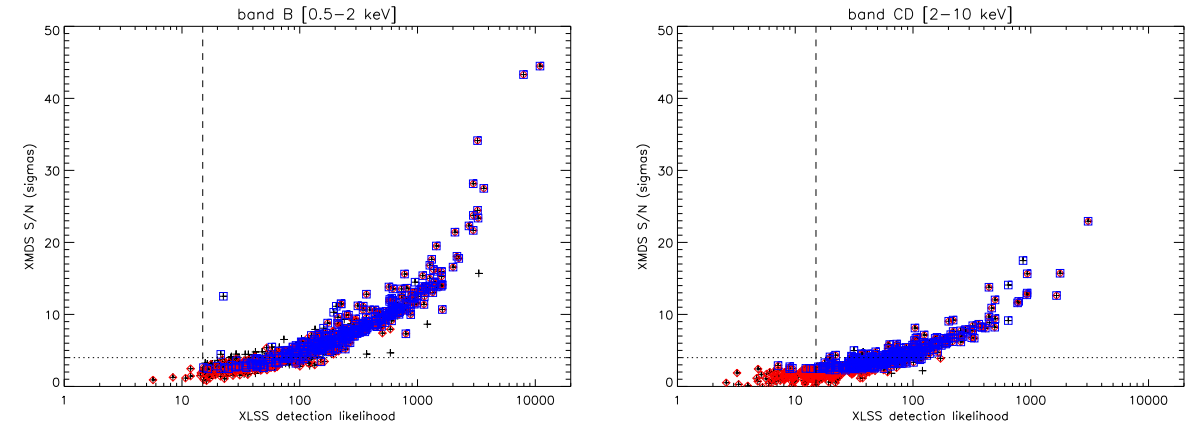

Fig. 16. Cross calibration between the `XLSS` (Xamin) detection likelihood and the XMDS signal to noise ratio. Left panel for the soft band, right panel for the hard band. The dashed vertical line indicates the `XLSS` acceptance thresholds of $ML > 15$ (remember that a source can be accepted if it is above the threshold *in at least one band* but not necessarily in all), while the dotted horizontal line shows the conventional level of 4σ. Crosses indicate all objects detected in the given band. A (red) diamond surrounds the sources detected above ML threshold *in both bands* in the `XLSS`, while a (blue) square surrounds those detected above chance proability threshold *in both bands* in the XMDS.

Acknowledgements. I acknowledge conversations with Laura Maraschi, who instigated the idea of capped probabilities, and thank Marzia Tajer for the collaboration in the inspection of the identification, and the idea of the plots in Fig. 13.

References

Baldi, A., Molendi, S., Comastri, A., et al. 2002, Ap.J, 564, 190
Bondi, M., et al. 2003, A&A, 403, 857
Chiappetti, L., et al. 2005, A&A, 439, 413 (Paper I)
Chiappetti, L., 2006a, XMM-LSS Internal Report N. 1-Mi (Report I)
Chiappetti, L., 2006b, XMM-LSS Internal Report N. 2-Mi (Report II)
Cohen, A. S., et al. 2003, Ap.J 591, 640
Le Fèvre, O., et al. 2004, A&A, 417, 839
Pierre, M., et al. 2007, MNRAS submitted
Polletta, M., et al. 2007, ArXiv Astrophysics e-prints, arXiv:astro-ph/0703255
Tajer, M., et al. 2007, ArXiv Astrophysics e-prints, arXiv:astro-ph/0703263

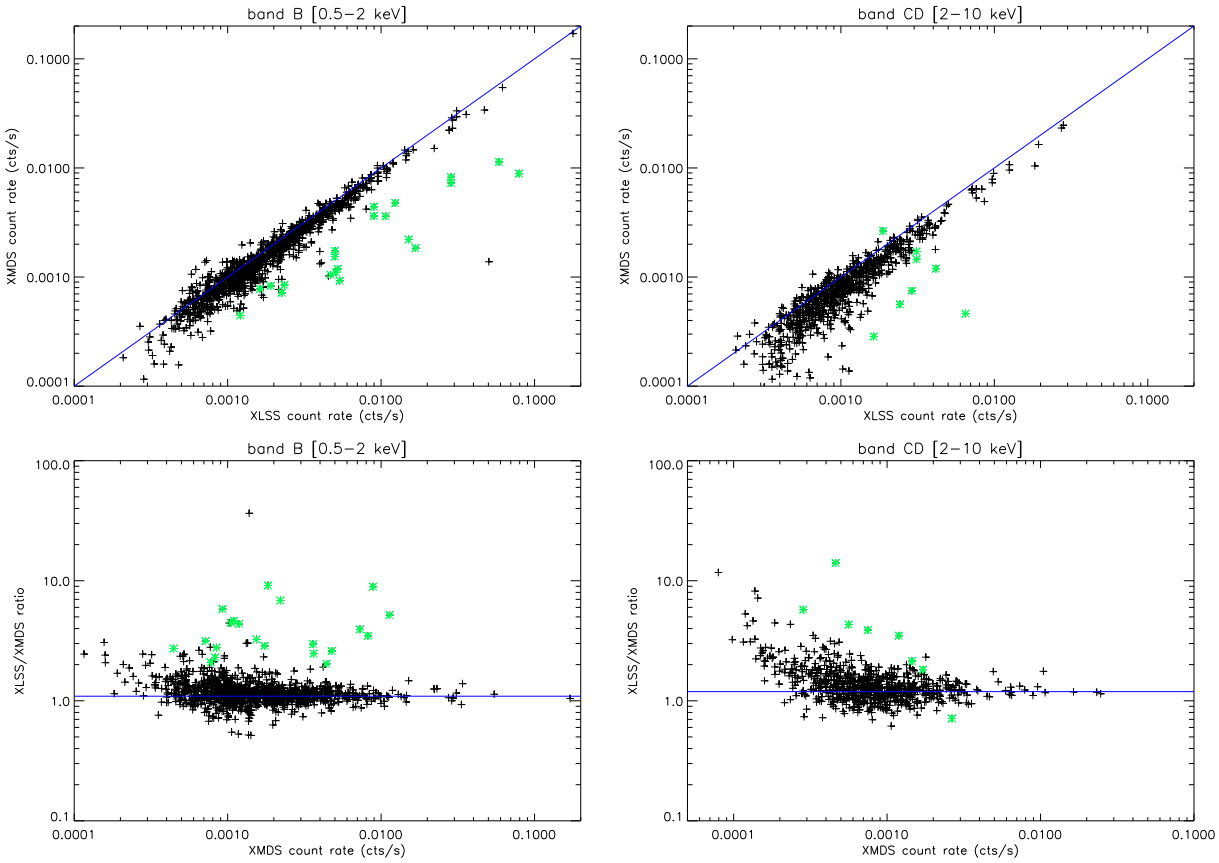

Fig. 17. Two alternate ways of comparing count rates. Top : The XMDS count rate vs the `XLSS` camera-merged count rate for band B (left panel) and band CD (right panel). (Black) crosses mark pointike sources detected in the band, while (green) asterisks correspond to extended sources. The diagonal solid line is a fiducial line corresponding to equal XMDS and `XLSS` rates. Bottom : The ratio of the `XLSS` camera-merged and XMDS count rates as function of the XMDS count rate for band B (left panel) and band CD (right panel). (Black) crosses mark pointike sources detected in the band, while (green) asterisks correspond to extended sources. The horizontal solid line is a fiducial line corresponding to the actual average ratio in the band (see text).

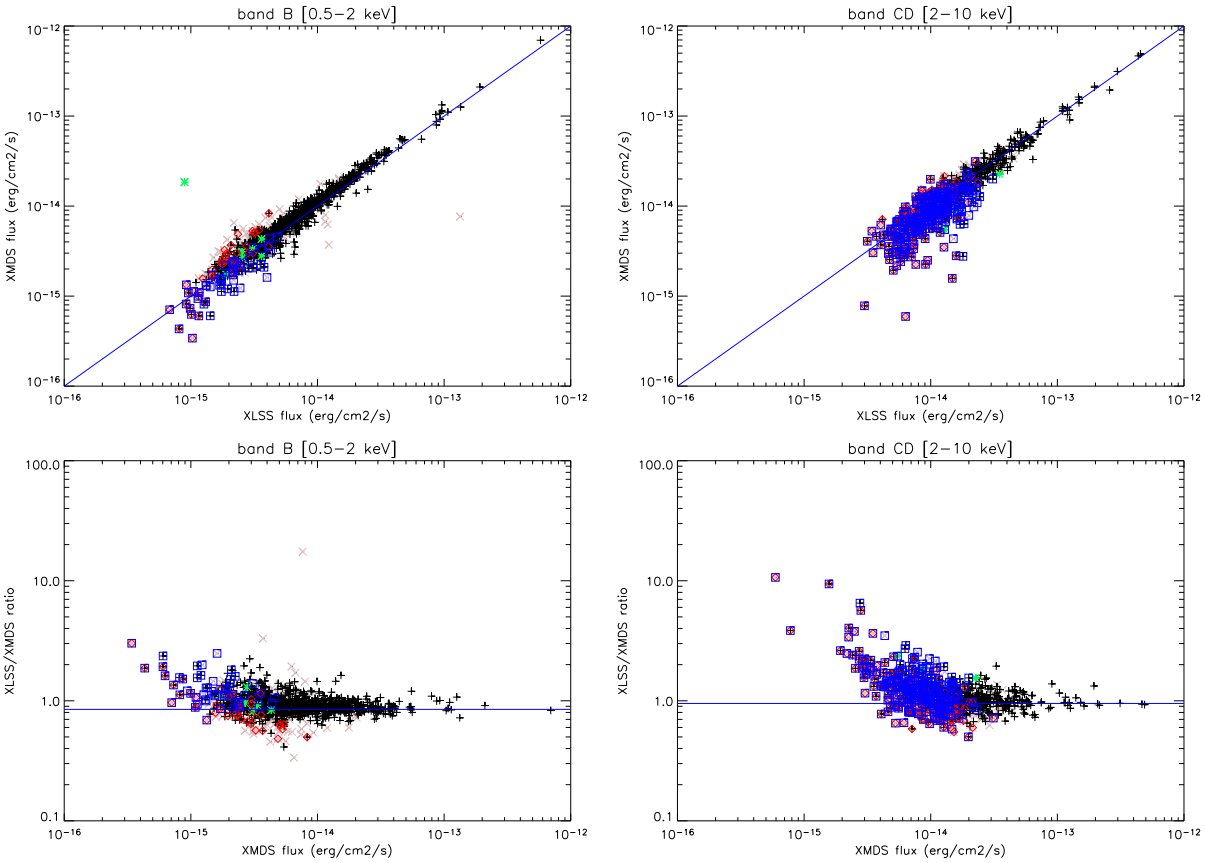

Fig. 18. Two alternate way of comparing fluxes. Top: The XMDS flux vs the `XLSS` flux for band B (left panel) and band CD (right panel). The diagonal solid line is a fiducial line corresponding to equal XMDS and `XLSS` fluxes. Bottom: The ratio of the `XLSS` and XMDS fluxes as function of the XMDS flux for band B (left panel) and band CD (right panel). The horizontal solid line is a fiducial line corresponding to the actual average ratio in the band (see text). Symbols common for both: The (black) crosses indicate pointlike sources which have a `fluxflag` of 0 or 1, the (pink) X those with a `fluxflag` of 2, i.e. where the MOS and pn fluxes differ by more than 50%. The (green) asterisk correspond to extended C2 sources for which the flux is computed from the pointlike rates (C1 sources have flux set to undefined and are not plotted). A (red) diamond surrounds the points with a poor `XLSS` likelihood $15 < ML < 20$ in the band. A (blue) square surrounds the points with a poor XMDS probability $p > 2 \times 10^{-4}$ in the band. Note this symbology is different from the one used in Fig. 16.