XMM-LSS
INTERNAL
REPORT n°10-Mi

May 23, 2012

Update to 2XLSS deep and 10ks catalogues

Fixing a bug in SXDS "chunks"

L.Chiappetti[1]

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

Abstract. I report on an update due to a bug in the "chunking" procedure of SXDS fields which affects the 2XLSSd deep catalogue (only the 40ks chunk i.e. field S01_40) and the 10ks 2XLSS catalogue (only fields S01 to S07)

Key words: LSS

1. Introduction

This document describes an update to the 2XLSSd and 2XLSS catalogues due to a bug discovered after the issue of Report IX in November 2011 (Chiappetti , 2011). The bug was detected while drafting the catalogue paper, and namely while computing the exposure table.

The bug affects only the curtailed exposures ("chunks") of the SXDS fields, which were processed before other chunks. The pn exposures were truncated earlier than they should have been. The affected fields are all 7 "subaru" fields (S01 to S07) in the 10ks 2XLSS catalogue, while in the deep 2XLSSd only the 40ks chunk of S01 (field named S01_40 and numbered 2901) was affected.

The correct chunking procedure was therefore applied to the affected field and XAMIN was promptly re-run. It was afterwards necessary to insert the corrected data into the database with minimal disruption to the previous work.

Although a limited number of fields and sources are affected, the update of the database, which was already in a final state, has been rather tricky, and this is the reason why this report has been written.

2. The 40ks field

I will describe in detail the corrective actions taken for 2XLSSd , where a single field is affected. I used this as a testbed for the procedure later applied also to the 7 10 ks chunks in 2XLSS (see 3).

I remind that the procedure leading to the generation of a catalogue passes through several steps:

1. two runs of XAMIN (one in the soft and one in the hard band), which for each field produce the respective FITS catalogue (or "cat"). Entries are identified by a numeric id (BOX_ID_SRC, whose values may repeat from one field to another.
2. the database *ingestion* in the 2 individual band tables takes the XAMIN parameters (plus the off-axis angles from a separate file). It keeps both the original pointlike and extended fit results (in particular the position), and applies a classification (PE and C1/C2 for the extended case). All entries are preserved but flagged spurious if pointlike and $ML < 15$. All entries are assigned a *sequence number* (seq) which is unique across fields.
3. the *band merging* generate a new table, where a source is classified as non-spurious if it is such in at least one band, and as extended if it is such and detected in one band only, or detected in two bands and extended in the soft one (i.e. E-, -E, EE, EP are extended, P-, -P, PP, PE are pointlike). The position and rates are taken according to the classification. Each merged source has its own unique seq (as well as a combined 6-digit id, for instance merged.seq=38741 resulting from the merging of b.seq=26232, b.id=103 and cd.seq=18246, cd.oid=78 has merged.id=103078)
4. some operations are done after band merging, like computation of positional errors and fluxes, and application of astrometric correction (the same for all kinds of coordinates). Some of these operation may exploit the correlation between the seq of soft, hard and merged entries.
5. the next step of *overlap removal* considers only sources flagged as non-spurious, and associates sources in overlapping pointings based on the astrometrically corrected coordinates. The seq of the sources surviving overlap removal are used as pointers in the X-ray catalogue (Xseq). The seq is the same, but there are less objects !

6. finally the preferred position is used for the multiwavelength identification, which also carries forward the same `Xseq`.

It is obvious that even adding a few pn events might alter the rates and the position, and therefore the spuriousness and classification. It is also obvious that XAMIN might not detect some old sources, or might detect a new ones. This means that the id and seq numbering can be altogether different from the very first step!

The goal of the update procedure is to track which updated sources are associated to pre-existing sources, so that one can *replace* the values (positions, fluxes etc.) keeping the same seq's for confirmed sources, assigning new unique seq's to new sources, and keeping track of the old "lost" (unconfirmed) sources.

2.1. Individual band ingestion

I have performed the ingestion of field 2901 in two semitemporary dedicated tables `apr12b` and `apr12cd`. They have to be compared with the data used for `2XLSSd` i.e. tables `jan11b` and `jan11cd`.

The overall number of sources is apparently similar (144 new soft vs 146 old, 135 new hard vs 132 old), but associating them is not immediate, because one has the choice of comparing the pointlike or the extended position, and has no guarantee the classification is remained the same. In the most favourable case (two soft C1 confirmed as such) the extended position has changed respectively by 1.7 and 5.4″. At the end I decided to make the association using the input X and Y coordinates in pixels, using a 3-pixel matching range.

This associates, in the soft band, 125 cases all distinct, of which one degrades from C2 extended to pointlike, and 1+5 change spuriousness. In the hard band one has 120 cases (120 old distinct with 119 new distinct), of which 1 pointlike becomes C2 and 4+6 change spuriousness.

One loses 17 soft and 12 hard sources (respectively 15 and 10 were spurious), and gains 15 soft and 16 hard sources (respectively 10 and 15 are spurious, and one hard is extended !).

There is an acceptable match in likelihood, and the MOS number of counts matches well when not too low, and with a reasonable scatter. The pn counts instead are (obviously) now 10% higher, scaling with counts.

So I generated for each band a semi-temporary *match table* which associates the old and new seq and id (id derives straight fron XAMIN while the new seq (re)starts from 1). They will be used in what follows.

2.2. Band merging

The standard 10″ band merging procedure is applied to the individual band tables, generating another semi-temporary merged table `apr12`. Of course this reshuffles and potentially reclassifies sources.

The nominal number of sources is rather similar (216 now instead of 218 in `jan11`). But using the distance between nominal merged positions might not be a wise choice. One has 196 matches within 6″ (196 within 10″, 173 within 2″). Of these 181 have the same classification, 3 are compatible (2 pointlike now `PP`, and one soft C1 now `EP` still C1), 7 are "degraded" or reversed (`EP` to `PP`, `P-` to `-P`, `PP` to `P-`, `PP` to `PE`). Of the 181, 169 have the same spuriousness, 2 are degraded and 10 promoted. Of the 10, 8 were and still are spurious, the other two swap spuriousness. The 191 are not distinct, there is one "crossed" couple of the so-called suspect (ambiguous soft-hard merging).

The majority of the old lost sources (26/29) or of the new ones (22/27) are spurious and single-band detections.

A better way of matching sources is to use the single-band match tables, and match on the matching id's in the individudal bands. This gives 59 cases of double-band detections, confirmed, 65 soft-only detections confirmed, 58 hard-only detections confirmed. Then there are 2 old double-band detections now detected only in one band, and 4 old single now detected in two bands.

In one case an old soft source matches a new hard one within 4.7″. In another cases the band merging considers separately `jan11.id=091000` and `apr12.id=089075` despite the extended position being within 1.7″, because soft `jan11b.id=91` matches `apr12b.id=89` but hard `apr12cd.id=75` matches `jan11cd.id=83` which is standalone.

If one constructs the merged *match table* from the individual band ones, one gets 188 matches, 30 old lost, and 28 new sources. The two special cases above fall automatically in the latter two categories.

2.3. Post band merging

The insertion of position errors and fluxes in the merged table is standard. The astrometric correction uses the pre-existing T004 offsets of field S01.

Fig. 1 compare the fluxes, which usually match quite well for non-spurious sources. There is one largish deviation (80%) in the soft band for `jan11.seq=38602` whose flux moves from 4.7×10^{-15} to 8.8×10^{-15}, possibly due to a pn gap. In the hard band there are six large deviations, of which three are not shown in the bottom panel (for the same `jan11.seq=38602` which is anyhow spurious in the hard band in both cases, and for `jan11.seq=38751, 38767` which are not detected in the hard band in the new case, but had anyhow rather poor likelihoods, 11 and 16, in old). `jan11.seq=38625` doubles its flux, is near a pn gap and spurious in both cases. `jan11.seq=38641` increases by 51% passing from spurious to non-spurious (ML 9 to 24), while `jan11.seq=38637` is the only non-spurious

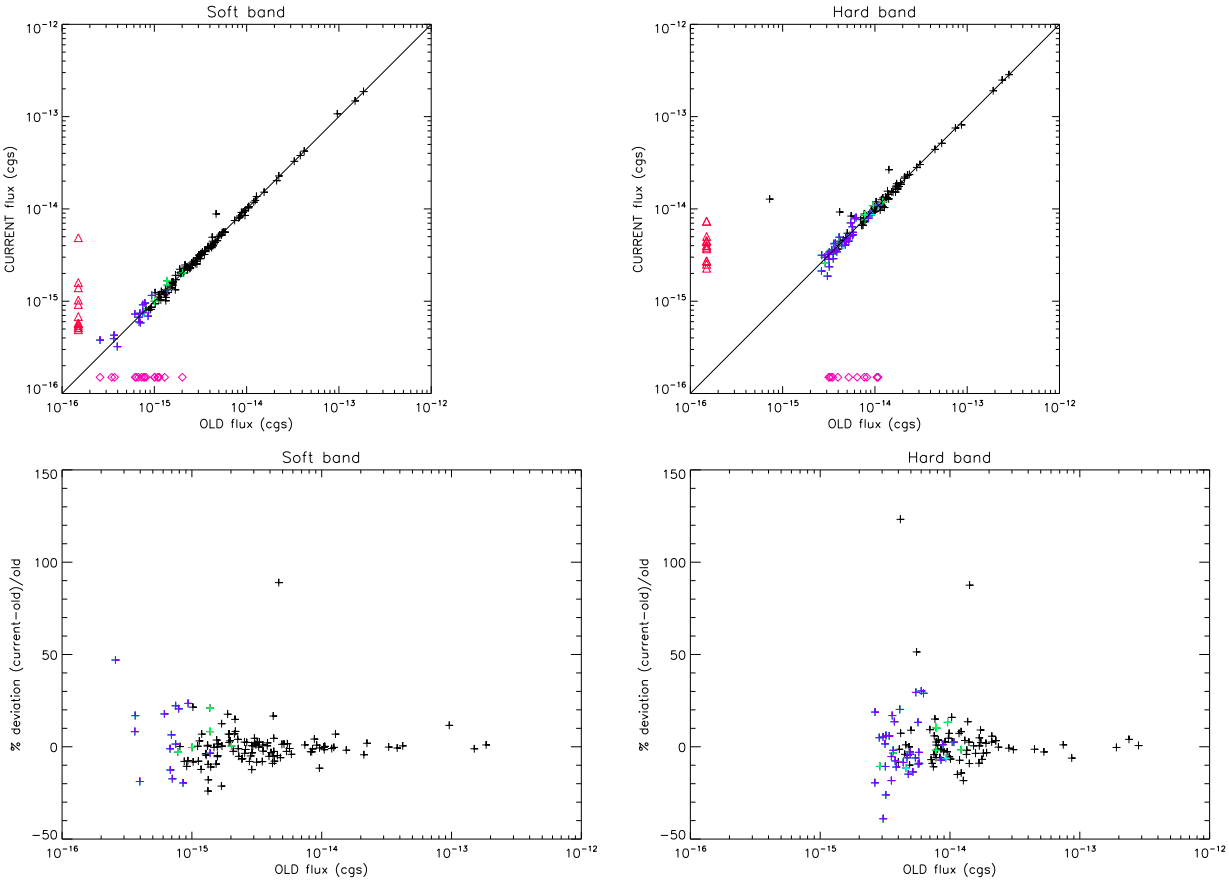

Fig. 1. Comparison of the old and new fluxes for field S01_40 (2901). Left column: soft band; right column: hard band. Top row: direct comparison of new vs old fluxes. Bottom row: percentage deviation of new flux w.r.t. old one. Black cross are sources detected in both old and new cases as non-spurious. Green, cyan and violet crosses are respectively spurious in old, new or both. Magenta diamonds and red triangles (at a conventional position) report the fluxes of sources detected only in old or new.

case, whose flux is enhanced by 88% with ML increasing from 52 to 161.

2.4. Overlap removal

In `2XLSSd`, only 1 of the 132 non-spurious sources on `jan11` was removed by the overlap removal procedure (`seq=38726`). A preliminary examination of the temporary match tables shows that 126 sources will be confirmed (and for these one has to repoint the seq number and replace positions, fluxes and other parameters), while 5 sources (`38622 38633 38670 38678 38734`) are "lost" and shall be removed.

However there are 6 non-spurious new sources which have to be examined for possible overlap removal. Just *for 6's sake* (Genesis 18:16-33) the overlap procedure has to be repeated afresh !

The old GCT (`glorlss11`) supporting `2XLSSd` had 131 overlap-free entries in field 2901. The new GCT will have 140. But these 140 will include only 123 sources already in `glorlss11`, 11 sources which were in `jan11` but were not considered for the catalogue because spurious, but which are now above the $ML > 15$ threshold, and 6 new sources. The previous 131 entries include the 123 confirmed, 5 "lost" sources (no counterpart in the new analysis) and 3 from the old catalogue which are not confirmed (2 are now spurious, and one was dropped in favour of an overlapping S07 source, and moreover is at an offaxis angle of 12.98′!).

Moreover there are sources in fields *other than 2901* which are no longer in the new GCT. Actually these are 11 S01 sources (the 80ks full exposure, field 2001, flagged nominally bad to give it a lower preference w.r.t. 2901, the 40ks one) which are now supersed by the entries in field 2901. For 9 of them in the old GCT the S01 source was preferred to the old 40ks counterpart, while now the S01_40 is preferred. The remaining 2 are part of the 6 new sources listed above.

2.5. Optical identification

As a prerequisite to the optical identification procedure, one has to make sure all possible counterparts in W1, SWIRE, UKIDSS and GALEX (of course none in D1, since D1 does not overlap SXDS; and only UDS for UKIDSS, since only UDS overlaps SXDS) around the new sources are loaded in the appropriate database tables and correlated. This exercise was run in cumulative way for the new source positions in the 40ks field (2901) and in the 7 10ks chunk fields.

After that, a new GCT (including pointers to the non-X-ray tables) is created from the new X-ray only GCT and the identification and ranking procedure is run. For simplicity, after the first two stages (initial population and X-ray placeholder record insertion) only the `field=2901` sources are kept in the GCT (this is equivalent to a range of `seq`s 38602-38823 and 39000-39027), since the identification procedure is independent for each field.

The resulting GCT is compared with the `field=2901` subset of the GCT underlying `2XLSSOPTd` i.e. `gloropt11`, with the following results

Case	new GCT	old GCT
Total entries	627	598
Excluding placeholders	532	501
Distinct X-ray sources	140	131
of which common	123	=
only in one case	new 17	lost 8

83 out of the 123 common sources have the same multiplicity (i.e. the same number of candidate counterpart sets). For 22 and 18 of the new sources, they have respectively more or less counterpart sets than `jan11`.

A more thorough examination of the individual counterpart sets (excluding the placeholder records)

X-ray source	counterparts	new GCT	old GCT
same	same	429	430
same	different	34	32
different	same	3	2
only in one	different	new 66	lost 37

The very few cases with the same counterpart set associated to different X-ray sources correspond to (usually marginal) counterparts of "lost" sources now associated to a somewhat displaced "new" source (`seq` above 39000).

The thirty-or-so cases with different counterpart sets are in about 75% of the cases marginal counterparts (usually rejected ones) which are either "lost" or "new". In 8 cases they are compatible counterpart sets (e.g. a WSUG and a WUG where the SWIRE counterpart is associated only in the old or new case), but they are also marginal (rejected or secondaries, i.e. rank -1 or 2) except one `rank=1` case.

For the vast majority of identical counterpart sets, in 88% of the cases (378) they have also the same rank. The remaining case are divided, in small numbers, about various possibilities (`rank` 0/1 now secondary or rejected or v.v., secondary now rejected or v.v. Otherwise said, of 123 common X-ray sources, 96 have the same *best* counterpart with the same rank, and 106 have the same best counterpart with compatible rank (0/1 or v.v.).

In conclusion the effect of the small positional displacement between old and new X-ray sources has little effect on identifications.

2.6. Corrective actions

One should make sure that the new sources (`apr12`) associated with the old ones (`jan11`) are assigned the **same** `seq`.

The sources which are altogether new in `apr12*` (inclusive of the single band tables) will receive a **new** `seq` **above 39000**.

Of course all the database columns which point to other source's `seq`, like `suspect` or `linkedto` shall be (re)numbered consistently.

After this one shall remove (from all physical tables and GCTs) *all* old entries pertaining to `field=2901` and insert in their stead the new ones. In addition the 11 S01 sources mentioned above have to be removed from the GCT `glorlss11` (but not from the physical tables). Their `Xseq` are: 17922 17929 17956 17981 18020 18068 18091 18099 18109 18186 and 18193.

Parallel corrective actions have to be performed for the GCT underlying `2XLSSd`, i.e. `glorlss11`, and for the one underlying `2XLSSOPTd`, i.e. `gloropt11`.

The band GCTs (`glorlssb11` and `glorlsscd11`) will be regenerated directly from the main GCT.

Once the update has been performed one has to recreate (update) all correlations tables depending on `jan11` with all other primary (D1, W1, SWIRE, UKIDSS and GALEX) tables, as well as wich additional ones (SIMBAD, NED, USNO, Stalin, Ueda, and the compilation by Olga), and propagate to `2XLSSd`.

Finally one has to update the data products (CFHT and SWIRE thumbnails around new sources). In particular in this circumstance I have loaded thumbnails (where existing) for all X-ray source positions (including those without a counterpart in the given band), and unconditionally replaced those for `field=2901`. Old ones have been retired offline. There have been some difficulties due to the long time elapsed since last time such data products were loaded (mainly little interface changes at CADC and IPAC), which have been overcome. In particular the SWIRE 70 and 160μm thumbnails for `field=2901` are linked to the retired version.

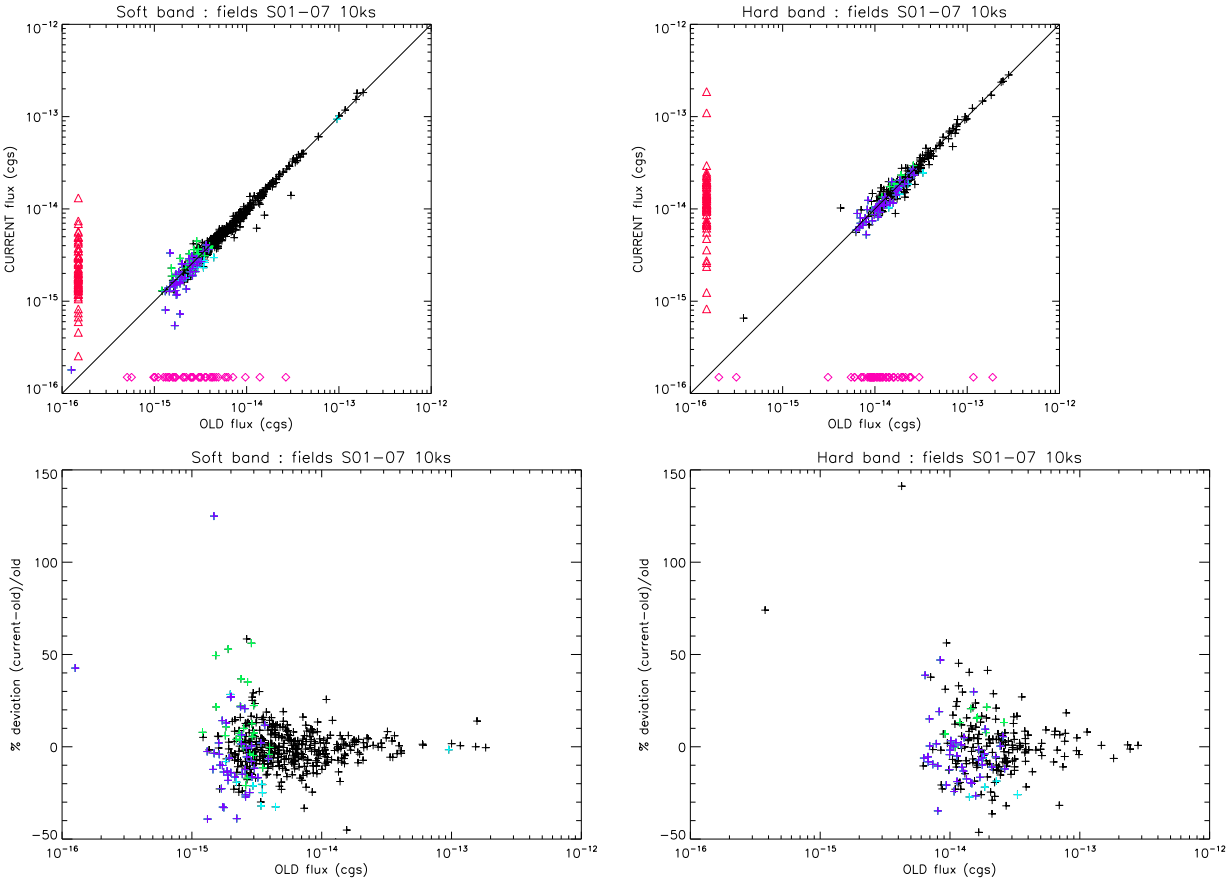

Fig. 2. Comparison of the old and new fluxes for fields S01-S07 (10ks). Panel arrangement and symbols like Fig. 1.

3. The 10ks fields

3.1. Individual band ingestion

The 7 10ks fields (still called S01-07 or numbered 2001-2007 in `2XLSS`) have been ingested in the same semitemporary tables `apr12b` and `apr12cd`. The 10ks data have however to be compared with the data used for `2XLSS` i.e. tables `may11b` and `may11cd`.

The overall number of sources is similar (534 new soft vs 507 old, 321 new hard vs 289 old), and so is the breakdown in individual fields.

As described in the corresponding subsection of Section 2 the old-new association was done using the input X and Y coordinates in pixels, using a 3-pixel matching range.

This associates, in the soft band, 469 cases, all distinct, with a very limited number of classification changes (two pointlike now C2, one C2 now pointlike). In the hard band there are 245 matches (245 old distinct with 244 new distinct) of which two old C2 are now pointlike. One loses 34 soft and 42 hard old detections, and gains 65 soft and 77 hard new ones. No further detailed statistics is provided.

Here too I generated for each band a semi-temporary *match table* which associates the old and new seq and id. They will be used in what follows.

3.2. Band merging

Again the standard 10″ band merging procedure is applied to the individual band tables, appending results to the semi-temporary merged table `apr12`.

The nominal number of sources is rather similar (668 now vs 635 in `may11`). Using distance as a criterion one has 541 matches within 6″ (435 within 2″ and 560 within 10″). 505 have the same classification, 491 have the same spuriosity. The majority of the old lost sources (71/84) or of the new ones (97/127) are spurious and single-band detection.

However here too the preferred way of matching sources is to use the single-band match tables, and match on the matching id's. This gives 155 double-band detections, confirmed, 285 soft-only detections confirmed, 76 hard-only confirmed, plus 2 old double-band detections now detected only in one band and 23 old single now detected in two bands.

Here too a merged *match table* is generated, with 541 matches, 94 old lost and 127 new sources in the seven SXDS 10ks chunks.

3.3. Post band merging

The insertion of position errors and fluxes in the merged table is standard. The astrometric correction uses the pre-existing T004 offsets of fields S01-S07.

The comparison of fluxes is presented in Fig. 2. Apart from the larger scatter this is rather similar to Fig. 1.

3.4. Overlap removal

Here too the simplest approach has been to repeat afresh the overlap removal procedure on a semitemporary new GCT. This new GCT will have 410 overlap-free entries in fields S01-07, while the old GCT (`glormay11`) had 384. 350 of such sources are in common, 27 were in `may11` but weren't considered for the catalogue, and 33 are new. The previous 384 entries include the 350 confirmed, 19 lost sources and 15 from the old catalogue which are not confirmed.

Moreover there are two sources in fields *not in SXDS* which are no longer in the new GCT, because the overlap removal procedure prefers a source in one of the S01-07 fields (they were 46936 in field B49 and 47243 in B52).

3.5. Optical identification

Here too a new GCT with pointers to the non-X-ray tables is created from the X-ray only GCT and the identification and ranking procedure is run as done for the 40ks field.

The resulting GCT is compared with the `field between 2001 and 2007` subset of the GCT underlying `2XLSSOPT` i.e. `gloroptmay11`, with the following results

Case	new GCT	old GCT
Total entries	1421	1312
Excluding placeholders	1388	1287
Distinct X-ray sources	410	384
of which common	350	=
only in one case	new 60	lost 34

233 out of the 350 common sources have the same multiplicity (i.e. the same number of candidate counterpart sets).

A more thorough examination of the individual counterpart sets (excluding the placeholder records)

X-ray source	counterparts	new GCT	old GCT
same	same	1043	1043
same	different	128	139
different	same	12	14
only in one	different	new 205	lost 91

The dozen or so cases with same counterpart set for different X-ray sources are (excluding the few placeholders or blank fields) one for an X-ray source preferred in the new GCT to an old one (51031 instead of 51493) by the overlap procedure; one counterpart of the lost 51305 which is now common with one of the 8 counterparts of the common source 51279 (which however moved by 2.4″). The remaining cases are (several) counterparts of 4 X-ray sources which are nominally new (39193, 39207, 39697 and 39442), which coincide with (most or all) the counterparts of 4 nominally lost X-ray sources (51003, 51042, 51505 and 51265). Considering however the small distance (0.58″, 1.25″, 0.58″ and 6.08″ however with a change of class from `PP` to `EP`) it seems wise to associate the old and new sources and *renumber them manually* with the old `seq`.

For the cases with different counterpart sets, 20 old and 16 new are ranked 0 or 1, and of them 5 are compatible (e.g. two WSUG with different G or a WG vs WSG).

For the cases with identical counterpart sets, 89% have the same rank, which goes up to 92% including compatible ranks (0/1 or v.v.).

In conclusion here too, the positional displacement between old and new positions has little effect on identifications.

3.6. Corrective actions

Here too the goal is to ensure that the new sources (`apr12`) associated with the old ones (`may11`), and all depending pointers, are numbered consistently (e.g. with the same `seq`). These `seq`s are historically numbers above 50937 (where `may11` by construction has `seq>=40000`).

The sources altogether new in `apr12*` will receive **a new seq above 39100**.

Then one shall remove (from all physical tables and GCTs) *all* old entries pertaining to `field between 2001 and 2007`, and insert in their stead the new ones. In addition the two sources 46936 and 47243 in fields B49 and B52 have to be removed from the GCTs These actions are done in parallel on `glormay11` (underlying `2XLSS`) and on `gloroptmay11` (underlying `2XLSSOPT`), while the band GCTs (`glormayb11` and `glormaycd11`) are regenerated afresh.

Finally one has to recreate (update) all correlation tables depending on `may11` with all other primary and additional tables, and propagate this to `2XLSS`.

No action are requested for data products, since these weren't loaded for `2XLSS`.

Instead the GCTs were updated for what the hidden column `deep` is concerned (this column provides a pointer from `2XLSS` into `2XLSSd`).

Acknowledgements. I acknowledge the work done by N.Clerc (who discovered the bug and promptly reran the XAMIN pipeline on affected data)

References

Chiappetti, L., 2011, XMM-LSS Internal Report N. 9-Mi (Report IX)