XMM-LSS INTERNAL REPORT n°17-Mi

October 15, 2014

Recent updates to XXL database

The XAMIN 3.3.2 sources

L.Chiappetti[1], L.Faccioli[2], and F.Pacaud[3]

[1] INAF, IASF Milano, via Bassini 15, I-20133 Milano, Italy
[2] Laboratoire AIM, CEA/ DSM/Irfu/SAp, CEA-Saclay, F-91191 Gif-sur-Yvette Cédex, France
[3] Argelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany

Abstract. I report on the recent update of the Milan database, which incorporates results from XAMIN 3.3.2 (curing the NaN likelihood issue for some bright sources), updating the parameters for 433 detections (or 132 catalogued, overlap-free, non-spurious sources).

Key words: XXL; data products

1. Introduction

The Milan DB has been originally populated (for what concerns X-ray data) using the results of the XAMIN 3.3 pipeline (Chiappetti, 2013a,b).

An improvement to such pipeline was announced in early 2014 and finalized after a teleconference (Faccioli et al., 2014a). A further report was posted on the wiki in March (Faccioli et al., 2014b). This latest version is known as XAMIN 3.4, and has never been released. It might not be released at all as a more advanced version is under development. However 3.4 fixes the main problem reported, i.e. the fact that in a few hundred of cases the *pointlike detection likelihood* assumes an undefined (IEEE-754 NaN[1]) value. Undefined values are equated to -1 during database ingestion, and therefore affected objects are considered spurious (likelihood below threshold of 15). Interest in the recovery of NaN-affected bright sources was expressed during an AGN teleconference in mid-March.

A comparison of the 3.3 database content with a preliminary list of XAMIN 3.4 sources (*not* in the format required for database ingestion) was made (and reported privately[2]) in April, and showed that a new release would contain new sources as well as sources with changed positions, the loss of some sources, and the overall change in source identification and numbering (`seq`). For these reasons co-existence of 3.3 sources with 3.4 sources (those affected by the NaN issue and cured) proved not easy to manage. Also the private release of partial lists was not useful because some post-processing steps done during database ingestion were missing.

Finally in the framework of last consortium meeting in Sesto, it was agreed to reprocess *all pointings used in the database* with a *descoped version* of XAMIN 3.4, hereafter known as XAMIN 3.3.2, which would cure the NaN problem, but provide data in the *identical format* used by version 3.3, and for *the same list of sources with same position and identification*, therefore making easy the correlation with and update to the existing database.

The coding and execution of XAMIN 3.3.2 was completed during the summer, and at end August results processed both in Lyon and in Bonn were made available. After an investigation about numeric differences between the Lyon and Bonn versions, the latter was used as input to update the current database tables. The latter process is the main topic of the present document.

This report is arranged as follows: section 2 provides some background information about the differences of XAMIN 3.3 and 3.3.2 and their implications for the database. Sections 3.1 to 3.4 describe the update procedure which ingested 3.3.2 data in working tables in the database; section 3.5 compares 3.3 and 3.3.2 data; section 3.6 describes the application of the overlap removal procedure to the working tables; section 3.7 presents the actual update when 3.3.2 data was *inserted into* the 3.3 tables. Follow-up operations on non-X-ray (multiwave) tables are described in 3.8, while the actual content of the updated tables is presented in the subsections of section 4.

2. Background information

2.1. NaN data in Milan DB

The IEEE-754 Floating Point standard representation allows undefined values called NaN (Not a Number). NaN values can be generated as a result of operations with undefined results. NaN values are supported by the FITS format in which XAMIN results are supplied for ingestion

[1] `http://en.wikipedia.org/wiki/NaN`
[2] LC internal reference: e-mail #94 8 Apr 2014

in the Milan database, but they are not supported by the `mysql` database engine.

Therefore any NaN value (whatever could be the reason for Xamin to produce it) is converted before ingestion in an undefined placeholder value, which for most database columns is value −1. In particular it is −1 for the various max likelihood (ML) columns.

In the database this has two main effects:

– objects with $ML < 15$ are flagged spurious, therefore NaN objects are also flagged that way (therefore ignored in the final catalogues);
– in the band merging process, sources with undefined values (which may occur because a given parameter is actually undefined, or also because a source is not detected in a given band) may be ignored. In particular one cannot assign the "best band" (`bandid=2` or `bandid=3`) from which the merged source inherits some parameters like position

The original `north33` table contained 196 sources flagged `bandid=0`, and `south33` contained 158. Of these 194 and 155 were flagged spurious and assigned zero (undefined) RA and Dec. The remaining 2 and 3 respectively were flagged extended (so by definition non-spurious) and for them the position was recovered. The spurious entries did not enter the overlap-free catalogues `XXLN` and `XXLS`.

2.2. Changes in Xamin *versions*

The main reason for the pointlike detection likelihood `PNT_DET_ML` assuming a NaN value in Xamin 3.3 was that the initial source flux used in the fit (estimated by Sextractor) was too high so Xamin 3.3, taking this as input, computed a negative background for the source and the fit silently crashed. This problem is fixed by the two variants Xamin 3.3.2a and 3.3.2b by testing for negative background and reducing the flux accordingly before starting the fit.

In addition 3.3.2b fixes two more problems of 3.3, not directly related to the `PNT_DET_ML`=NaN problem but that in principle required solution:

– An improper resetting of the detection likelihood when it reached high values
– The fact that the center used to compute the off-axis angle was not the center of the 13′ detection Xamin radius (the difference is a few arcsec).

At the end we decided to use version 3.3.2a, as the changes in the off-axis angle might have side effects on the overlap removal procedure for *all* sources. Also the likelihood reset might modify (slightly) the fit results for many non-NaN sources. Both are considered at this stage unwanted updates. We however decided to accept *all* side-effect changes to *any* likelihood (not just the pointlike detection one) implied by 3.3.2a (the "reason codes" described in 3.5 and 3.7).

3. The update procedure

During the earlier tests some numeric differences were encountered between Xamin 3.3.2 run in Lyon and in Bonn (where the original Xamin 3.3 was run).

A sample test showed that the number of sources, the sequence number of sources, and some parameters were *identical* between 3.3.2 and 3.3; most other parameters for most sources showed *numeric differences* well within what possible for a 32-bit floating point number generated on different systems or machines (i.e. 6th decimal digit). There were of course *significant differences* for a number of parameters for NaN-affected sources as well as for a few other sources.

At the end we converged on the following procedure, which allowed to *re-use* most of the existing scripts, as well as to mimic the complete ingestion procedure, and to *assess the time needed for a full ingestion*, merging and postprocessing.

3.1. Material for update

What we finally used as input was a tar file with the Xamin FITS catalogues produced by Xamin 3.3.2a in Bonn for all pointings available there (615 in the soft band, 611 in the hard band, for a total of 616 distinct pointings).

Some of these pointings were excluded by purpose at the time of 3.3 ingestion because flagged as too bad (but there were better re-observations) or for historical reasons (for the SXDS fields S01, S03, S04 only the "combo" XXLn998-01z, XXLn998-03z and XXLn998-04z were traditionally used and the individual "chunks" (e.g. XXLn998-01a and XXLn998-01b) ignored).

So a list of the same 601 pointings already in the 3.3 database (274 in `north33` and 327 in `south33`) was compiled, and only such pointings were used for ingestion.

The relevant FITS catalogues were renamed according to the naming convention used for 3.3, and stored in a separate `v3.3.2` directory. The format of the FITS catalogue was otherwise *perfectly compatible* with version 3.3.

3.2. Ingestion

Ingestion of the Xamin 3.3.2 soft and hard band data was done by the standard ingestion script (without detailed inspection of the log files, i.e. in *unassisted mode*), reading from the `v3.3.2` directory and writing into clone database tables (`north332b`, `north332cd`, `south332b`, `south332cd`).

It is important to process the pointings in the same order in which they were processed (at various dates since they were supplied in up to 7 different epochs) for version 3.3. This should ensure that sources in the same field and with the same `id` receive the same `seq`. De facto this occurs in the south, while in the north 30 pointings received originally a different seq. This occurs because the

3.3 data for some pointings were supplied twice, with the newer one replacing the older one (and the released `seq` numbers not being re-used). It is enough to shift the 3.3.2 `seq` forward by the unused values, in order to align also the north data by `seq`.

Database ingestion (inclusive of 4 minutes for tar file retrieval and unpacking) in unassisted mode took just 11 minutes.

3.3. Band merging

Band merging creates the clone tables `north332` and `south332`, using the standard merging script. Here too processing shall be done in the same pointing order used for version 3.3, resulting in the same `seq` with the same provision of "shifting forward" a few northern pointings.

Band merging in unassisted mode took just 5 minutes.

3.4. Post-processing

The post-processing steps following band merging include: divorce of ambiguous mergings, generation of derived parameters (fluxes and position errors) and astrometric correction.

This used either the standard scripts, or some well codified manual operations (including the repetition of a few manual edits done for version 3.3). The astrometric correction required to assemble together the files with the correction coefficients used for 3.3 at various epochs (keeping the later values for a few pointings which were processed twice).

All this took a total of 56 minutes, where about half of the time was spent in manual operations and preparations. The most intensive processing time was 16 minutes for derived parameters (database indices not optimized ?), while astrometric corrections took only 2 minutes (this time is underestimated w.r.t. real life because it does not include correction coefficient computation … *existing* coefficients were just *re-used*).

3.5. Comparison

More than a couple of days was spent comparing the 3.3.2 band and merged tables with the pre-existing 3.3 tables (which are *no longer available* to database users, although they have been backed up).

I developed some `mysql` stored procedures to compare two `mysql` tables, reporting the number of entries (rows) for which *each* column differs by a predefined percentage (the procedures do not support non-numeric columns, and fail when the value in the older of the two tables is zero, but can be run in debug mode to inspect all deviations).

I ran the stored procedures on the individual band and merged tables in both northern and southern areas. Old (version 3.3) and new (3.3.2) tables are joined on `seq`. As expected, a number of parameters (like identifiers, exposure times, positions and, by construction, off-axis angle) are *identical.* The bad field flag, which was undefined in the 3.3.2 tables, has simply been forcefully inherited from 3.3.

One remains with a limited number of significant changes (where "significant" has been set by trial and error to a deviation of more than 5 % in absolute value; and "limited" means a few tens up to less than two hundred cases) in other parameters. These include the *three likelihoods* (pointlike detection, extended detection and extension) as well as counts, rates, background level, extent, position error and also classification parameters like pointlike vs extended, C1 vs C2, spurious vs non-spurious.

In the case of band merged tables also position sometimes change, but this is obviously an effect of either the solution of the NaN issue (change from zero-assigned to defined) or of a change in classification (the merged position is either the pointlike or the extended position, so if classification changes it might be "repointed").

What one would like to know is which changes are *native* to the changes in XAMIN 3.3.2, and which are just *consequences* of other changes. For instance the change in position error is obviously a consequence of changes in rates, since it is computed by tabulation from rates and off-axis angle. Otherwise said. one needs to find a *recipe* to identify sources to be updated.

An ad-hoc analysis was run, where changes in columns were flagged as none, significant (above 5%) or any (less, usually much less, than 5%). Likelihood changes had also special flags (when the *old* one was NaN, and also when it was zero, which sometimes occurred for extension likelihood when the detection likelihoods were NaN).

It can be clearly seen that (significant) changes in "soft info" or "hard info" (i.e. counts, rates, etc. in the soft or hard band) clearly depends on changes in likelihood.

Therefore at the end we defined five different "reasons", to be applied to the individual band tables:

- 1: pointlike detection likelihood has changed, starting from NaN
- 2: extended detection likelihood has changed, starting from NaN
- 3: $>5\%$ variation in pointlike detection likelihood
- 4: $>5\%$ variation in extended detection likelihood
- 5: $>5\%$ variation in extension likelihood

Reasons 1 and 2 can be called "NaN reasons".

For `north332`b we have 144 cases with "reason 1" plus some of the other reasons, 6 cases with "reason 2" (and 4), and 3 cases with some of the other non-NaN reasons.

For `north332`cd all 77 cases have "reason 1" and some of the other.

For `south332`b we have 160 cases with "reason 1" plus some of the other reasons, 4 cases with "reason 2" (and 4), and 2 cases with some of the other non-NaN reasons.

For `south333`cd there are 37 cases with "reason 1" plus some of the other reasons, 1 case with "reason 2" (and 4), and 2 cases with some of the other non-NaN reasons.

For the band-merged tables it makes no sense to make a direct a-priori comparison between 3.3 and 3.3.2 versions (also because here -1 can flag both a likelihood which was/is NaN in the individual band, as well as a likelihood - or any other parameter - which is undefined in a band because there is no detection in such band), but it is simpler to inherit the individual band flagging. A merged source is considered *affected* by a "likelihood problem" if *at least one* of its soft or hard component is affected by a reason 1 to 5.

In fact *all* changes in the merged tables reflect a change in one or both of the individual band tables, while there are a limited number of changes in the invididual band tables which are irrelevant. For instance `south332` source 201698 has a classification `EP` (reclassified extended because of the soft band), therefore the merged parameters are taken from the extended fit, and if there are NaN-related changes in a single band table in the pointlike fit, these do not find their way in the merged table.

Note instead that NaN-related changes in individual band tables may have more or less expected side effects in the merged table. In most cases to be updated, the "spuriosity" will change (old NaN likelihood were equated to -1 i.e. below threshold, and now can be > 15). In a limited number of cases the classification might change from pointlike to extended or v.v. In these cases also the position might change (it could e.g simply "repoint" from an unchanged pointlike to an unchanged extended position).

The situation for the overlap-free catalogues can be more complex and is described in the next section and in section 4.2.

3.6. Overlap removal

The overlap removal procedure is the first step towards the creation of an overlap-free X-ray *catalogue* and it starts generating a GCT (Generalized Correlation Table) which links the entries in a band merged table and its two associated soft and hard band tables. The simplest procedure to cope with XAMIN 3.3.2 data has been to generate two temporary GCTs (called `xxln332` and `xxls332`, and apply to them the standard procedure and overlap removal script. The standard procedure requires some manual preparation step, a dry run of the script, the actual execution of the script and some post processing manual steps (like assigning unique catalog names to ambiguous cases, associating `2XLSS` pointers (northern area) and generating GCTs for single bands.

The new GCTs will of course have more entries than the old GCTs (associated to catalogues `XXLN` and `XXLS`), both before and after the overlap removal, because of the new non-spurious sources (recovered NaN-affected sources). Since the new sources are interleaved with old ones, the sequence number in the GCTs will not match (the new and old GCTs can be joined on the sequence number in the primary member table, e.g. on `glorxxln.north33=xxln332.north332` which is the same as `north33.seq=north332.seq`).

The overall procedure required some 3.5 hours, of which only 30 min are taken by the actual overlap removal scripts.

At the end of this step, although no actual update to the database has yet been performed, one can already predict the effect on the catalogues:

- the majority (14127 in the north and 11854 in the south) of the sources will be unchanged
- a limited number (41 in the north and 34 in the south) of *new sources* will appear (former NaN-affected now recovered, with a non-spurious ML and surviving the overlap removal)
- a very small number of sources (7 in the north and 9 in the south) will be lost (and are listed in 4.2). In most cases they will be simply *replaced* by one of the new sources which will have a more favourable condition (better `badfield` flag or smaller off-axis angle) during the overlap removal. One case in the north will be replaced by a pre-existing source (old source 216257 removed in favour of new 216328, in turn removed in favour of pre-existing 226938). Another case in the south (211385) will instead disappear totally. This source was formerly classified `-E`, but is now classified `-P` and spurious. It is a non-NaN case for which both the (hard) pointlike detection likelihood and extesion likelihood change. The former changes from 13 to 2 (remaining under the spurious threshold), and the latter moves just from 15.11 to 14.98, which is however enough to fail the C2 recipe, making the classification as pointlike to prevail.

3.7. Actual database update

The actual database update has been split in three or four steps as follows:

3.7.1. Single band table update

First of all one inserts a new flag column (`v332`) in the 3.3.2 band tables (e.g. `north332b` etc.) and assigns them the 5-char coded reason (e.g. 10000 means reason 1, 2000 reason 2, 12000 reasons 1 and 2 etc.). This flag can be dealt with as a 5-char string or as a 5-digit number.

A new ad-hoc script (`updateband332.csh`) will then update the publicly advertised tables (i.e. `north33b`, `north33cd`, `south33b` and `south33cd`) from the 3.3.2 tables, for the objects having a *reason code* above an user-supplied threshold. *We decided to use a threshold of 1, i.e. update all objects for which any of the likelihood changed,* *rather than a threshold of 10000, i.e. confining only to pointlike detection NaN cases.*

For the objects where the reason code flag in the 3.3.2 table is above threshold, we update the entry in the 3.3 table as follows:

– a *simplified* `v332` flag will be set to 1 if the source is updated and remains 0 for the majority of unmodified sources
– the detection likelihoods as extended and pointlike, the extension likelihood, the core radius, the position error, the MOS and pn counts, rates, background and `pixdev` for both extended and pointlike cases, the `class`, C1/C2 and `spurious` flags are copied from 3.3.2 into 3.3

Note that the original 3.3 tables were backed up off-line before applying the update.

3.7.2. Merged band table update

Here too first of all one inserts a 5-char coded reason flag in the 3.3.2 band merged table (`north332` or `south332`) just adding up the numeric representation of the `v332` flag in the soft and hard band tables. So a source detected (and affected) only in one band will get the reason code of such band, a source detected in both bands and affected only in one will receive the reason code of the latter, and a source detected and affected in both will receive the sum of the reason codes.

The latter case (code 20600 sum of two 10300, reasons 1 and 3) is very rare (3 cases in the north and 1 in the south). Most affected cases (220 in the north and 196 in the south) are NaN-affected (code $\geq$ 10000), and just 9 in the north and 9 in the south are affected by non-NaN likelihood issues.

The update of the `north33` and `south33` merged tables is handled by the same script `updateband332.csh` contextually with the band tables.

For the objects where the reason code flag in the 3.3.2 table is above threshold, we update the entry in the 3.3 table as follows:

– a *simplified* `v332` flag will be set to 1 if the source is updated and remains 0 for the majority of unmodified sources
– the detection likelihood, extension likelihood, the core radius, the all-camera counts, rates, fluxes and flux flags and the hidden coordinates in both bands, the merged position (raw and astrometrically corrected), the `class`, `reclass`, C1/C2, `extended` and `spurious` flags, the `bandid` and `maxdist` are copied from 3.3.2 into 3.3

Here too the original 3.3 tables were backed up off-line before applying the update.

It took a bit more of 2 hours to develop the script, and some 14 min to run them, covering all physical table updates.

3.7.3. GCT update

The update of the GCTs `glorxxln` and `glorxxls` used for `XXLN` and `XXLS` from the new GCTs `xxln332` and `xxls332` occurs via a new simple script `updateGCT332.csh` in a straighforward way: entries which are new in the new GCT are appended to the old table, entries which are not present any more in the new GCT are removed from the old table.

It took 4 minutes to run the scripts, and some additional 19 minutes to perform some mostly manual post-processing.

Post-processing includes the propagation of the `2XLSS` flag in the northern GCT, and the regeneration of the single band GCTs (using the standard procedure and verification steps).

3.8. Follow-up operations

When the above operations were completed, a *fast track release* of the updated tables was advertised to database users, so that they could start working immediately on X-ray data, although the overall (multiwave) database was still unstable (the *following* operations were not executed).

Updates to non-X-ray tables (3.8.1) and correlation tables (3.8.2) were run later, without advertising. At this stage only the *multiwave catalogues* and data products remained unstable. With the release of this report, also all following operations were completed, and the database is now self-consistent.

3.8.1. Update non-X-ray tables

X-ray sources updated (flagged with `v332=1`) may change their (band merged) position as described in 3.5. Since non-X-ray tables are constructed as 10″ subsets around X-ray position, one shall assess whether the position change is relevant. While for sources which mantain the same classification and `bandid` (15 `north33` and 18 `south33` i.e. typically those detected in a single band), the distance difference is $< 0.0044''$ (numeric), for sources which mantain the same classification but change `bandid` (43 `north33` and 53 `south33`) position may change from 0.12 to 9.0″ (actually, it changes *exactly by the entire* `maxdist` !).

This means one has to reload all non-X-ray data.

The operation is different for the tables originally generated by me and advertised in the database (referred later as Lucio's) and the tables extracted by the homogenized files supplied by Sotiria Fotopoulou (and not yet advertised). See Chiappetti (2014a) for more details.

Processing of Sotiria's subset is easier (i.e. faster), because the bulk data are already resident in Milan. One has just to adapt the original scripts to skip the source density computation. Three scripts are necessary, one for the bulk of northern data, one for the southern data (the table names have an `s` prefix for internal database reasons), and a patched one for northern IRAC data (the original files were incorrect and replaced by new ones with some little naming differences).

This involves processing 14 northern and 6 southern database tables, for a processing time of 73 minutes (plus the time to edit the scripts).

Processing of Lucio's tables required first to collect all ad-hoc scripts used to retrieve the data, and to assess the edits required. Essentially there are 2 or 3 sorts of tables: those for which the 10″ subset is derived locally in Milan from a bulk data file already stored here, those for which the subset is derived locally from a data file stored elsewhere, and those for which the subset is derived using facilities of a remote database (e.g. IPAC Gator, NASA CasJobs, Edinburgh WSA) uploading there an X-ray source list.

The procedure involved 11 tables for the north, and 7 tables for the south (two of them, i.e. `galexgr6` and `wise` are common to N and S in my arrangement, but have to be processed twice with input X-ray sources of the two areas). In addition there are 4 "external" tables (NED, SIMBAD, USNO and Marseille) which are relevant to both areas. Some of the tables required no operation because the Milan DB already stores a bulk dataset (OM-SUSS and Marseille), or because they were populated taking all objects around the pointing centres (SIMBAD, USNO). In addition tables for SSDF V8 had been abandoned in favour of SSDF V9 long ago and weren't updated.

This leaves us with 17 tables, of which 4 using local bulk data sets, and the rest requiring access to remote datasets. Fortunately the arrangement on the remote site has not changed since last retrieval!

For the remote cases requiring upload of an X-ray source list, this was generated only for the `v332=1` sources. For cases requiring correlation run locally, instead of doing it on the entire `north33` or `south33`, this was done on a `VIEW` containing only `v332=1` sources. This should save some computing time. In some cases the retrieval was complicated by the fact the remote archive is arranged *per tile*, so a list of affected tiles had to be constructed in advance.

Only tables present in the database were updated from the very same release from which they were created; new releases which are know to exist (e.g. OM-SUSS, DECam, GALEX) were not considered as they will be part of the newer Sotiria's release foreseen for end 2014 as agreed in Sesto.

The pure processing time (including manual logging) for Lucio's tables amounted to 8 and half hours, but was preceded by more than 7 hours to recollect all scripts to be used. So in fact this took more than 2 working days.

At the end of this process all non-X-ray tables listed in Chiappetti (2014a) are updated to contain all objects within 10″ of old or new X-ray positions. The additions varied from a few tens objects to about 600 for the deeper tables (e.g. CFHTLS or DECam).

3.8.2. Update correlations

The next step has been to update the correlation tables between the updated X-ray tables (new positions) and the updated non-X-ray tables (new objects). These are customarily done within 6″, with a few exceptions. In additions in some unfrequent cases there were correlation tables between couples on non-X-ray tables, which had to be updated as well.

For Lucio's tables most correlations are done by the `correlate-superfast.csh` script, which generates the correlation table, handles registration, and requires logging to a log file. The exact calling sequence of the script for the various table combinations was recollected from the original journal files at the time the scripts for generation of the non-X-ray tables were recollected.

This steps includes also correlation with the GAMA related tables.

The processing time (including manual logging and some verification) amounted to 4 hours and 21 minutes.

For Sotiria's tables correlations are done by the `correlate-special.csh` script, which generates the correlation table, without registration, nor logging. The original calling sequence was simply repeated.

The processing was completed in 11 minutes.

3.8.3. Update multiwave catalogues

There is only one set of "true" multiwave catalogues (`XXLNOPT` and `XXLSOPT`) advertised as such in the database, and depending on their own GCTs `glorxon` and `glorxos`, which have as member tables the older of "Lucio's tables", as recalled in (Chiappetti, 2014a) (hereafter Report XIV).

In addition, as also described in Report XIV, there is one set of GCTs `testxon` and `testxos`, which have as member tables "Sotiria's tables" as well as all advertised "Lucio's tables". Counterpart sets in these GCTs are *ranked* (`2XLSS`-style), although the ranking is done only on "Sotiria's tables". These GCTs are used as "fake" catalogue `TESTXON` and `TESTXOS` in the thumbnail visualization tool described in (Chiappetti, 2014b) (hereafter Report XVI).

Similar to what done for the previous cases, the simplest way to proceed has been to generate an altogether new set of GCTs, starting from the *updated GCTs* glorxxln and glorxxls after the 3.3.2 edits described in section 3.7.3.

At the end of this step the old multiwave GCTs will be retired (and implicitly backed up) just renaming them, and the new multiwave GCTs will be renamed to the "official name", so that the catalogues (XXLNOPT, TESTXON, etc.) will automagically point to the updated GCTs.

For Lucio's catalogues the temporary GCTs were called glorxon332 and glorxos332. They have the same member list as the original GCTs and are created and populated using the same sequence of scripts used for XXLNOPT or XXLSOPT.

The resulting GCTs have 135 (north) or 80 (south) more entries than the original version. The internal sequence numbering is different due to the presence of new or lost X-ray sources. The difference in content will be described in section 4.4.1.

The excution of the scripts, once recollected, took about 2 hours.

For Sotiria's catalogues the temporary GCTs were called testxon332 and testxos332. They have the complete member list as the original GCTs and are created as populated using the same sequence of scripts, the same configuration files and the same correlation radii described in Report XIV.

In the north one has 71906 entries (154 more than before) of which 58830 use counterparts in Sotiria's tables and the rest are "old only" (as described in Report XIV). For the south the total is 58795 entries (150 more than before) of which 51607 use counterparts in Sotiria's tables.

The procedure includes computation of probabilities, pre-ranking and ranking and identification of ambiguities as described in Report XIV. Some statistics is provided in section 4.4.2.

The execution of the scripts, once recollected, took 6 hours and 20 minutes.

4. The updated database

4.1. Physical tables

The single band and band merged tables (north33b, north33cd, south33b, south33cd, north33 and south33) contain a new column, the boolean flag v332. It is possible that for some time the column be hidden (not advertised), but if you include in your query the condition v332=1 you will be able to filter only the updated entries. Conversely v332=0 gets the majority of unchanged entries. Remember there are no new entries (only unchanged or modified).

The following table gives a count of updated sources. Not all of them, despite they now have a non-NaN likelihood, qualify as non-spurious (i.e. $ML > 15$).

Table	v332=1	and spurious=0	still NaN
north33	228	115	26
north33b	153	105	29
north33cd	77	11	2
south33	205	127	33
south33b	166	113	36
south33cd	40	11	4

Note that XAMIN 3.3.2 did not resolve *all* NaN cases. The last column in the above tables list the survived ones (for the single band tables they are the cases strictly with detlik_pnt=-1, for merged tables with bandid=0). These residual NaN are all flagged v332=0 (i.e. unchanged) with one exception.

This is south33.seq = 210060 i.e. south33cd.seq = 204837, that is hard source with id=86 in field XXLs070-07. In this case XAMIN 3.3.2 solved a NaN for the *extended* detection likelihood (reason code 2040), but since the source remains classified pointlike and spurious this is irrelevant.

In north33, the updated soft-only pointlike sources (98, 52 non-spurious) cover a flux range (non-spurious only) of 1.7×10^{-15} to 3.9×10^{-14} cgs; the updated hard-only pointlike are 66 of which just 3 non-spurious, and the latter cover the flux range from 8.1×10^{-14} to 1.7×10^{-13} cgs; pointlike sources detected in both bands are 55 (51 non-spurious) covering a range of soft flux from 2.3×10^{-15} to 4.7×10^{-13} cgs and of hard flux from 8.3×10^{-15} to 8.4×10^{-13} cgs. In addition there are 9 extended sources (of which 3 hard-only).

In south33, the updated soft-only pointlike sources (95, 48 non-spurious) cover a flux range (non-spurious only) of 2.0×10^{-15} to 4.7×10^{-14} cgs; the updated hard-only pointlike are 27 of which just 3 non-spurious, and the latter cover the flux range from 3.6×10^{-14} to 3.2×10^{-13} cgs; pointlike sources detected in both bands are 72 (65 non-spurious) covering a range of soft flux from 2.4×10^{-15} to 1.0×10^{-12} cgs and of hard flux from 3.8×10^{-16} to 2.0×10^{-12} cgs. In addition there are 11 extended sources (of which 2 hard-only).

In particular, concerning the extended/pointlike classification, this changed for 9 north33 sources (in all cases from pointlike to extended, 5 soft-only, 3 hard-only and one PP becoming EP. In south33 there are 11 changes, one in direction extended to pointlike (the already mentioned 211385), the other all from pointlike to extended, 5 soft-only, 2 hard-only, 2 PP becoming EP and one PE becoming EE. All changed cases are listed below.

Change	Affected sources
north33	
P- to E- C1	223528 215150 215183
P- to E- C2	216328 212739
-P to -E C1	200836 200837
-P to -E C2	200848
PP to EP C1	219774
south33	
P- to E- C2	211691 219168
E- to P-	211385
-P to -E C1	223319 223335 223631
-P to -E C2	210038 218097
PP to EP C1	223334
PP to EP C2	220410
PE to EE C1	206598

4.2. X-ray catalogues

The XXLN* and XXLS* X-ray catalogues are VIEWs and therefore got automagically updated when the underlying GCTs were updated. We remind that catalogues contain only non-spurious sources which survive the overlap removal procedure.

De facto this means that, if one is interested in knowing which catalogues sources are affected by the Xamin 3.3.2 update, one has to proceed as follows:

- tick XXLN or XXLS to select the view
- tick also on the "Show member tables (and data products!) also" tick box which appears at the very bottom of the table list
- use the Advanced query interface
- add to the selection condition: `and north33.v332=1` for XXLN, or `and south33.v332=1`for XXLS, if one wants to select the updated entries only
  (the v332 member table column is currently hidden but can be called by name)
- *or* add column north33.v332 or south33.v332 respectively to the output column list, if you want to get an arbitrary list of sources and know which was or wasn't updated

Such queries will return a total of 61 "updated" sources for XXLN or 71 for XXLS.

However this does not tell whether the source was already present in XXLN or XXLS (and has simply been changed in some values) or whether it is new. Note that it is also possible that some sources present in XXLN or XXLS before 23 September 2014 have been removed in favour of new sources more favourably handled by the overlap removal procedure.

We provide below the list of new (41 XXLN and 34 XXLS) and lost (7 XXLN and 9 XXLS) sources, and for the complement of sources which were just updated. However be informed that the latter sources can be located using the following "hidden" condition: for the 20 XXLN modified sources use condition `glorxxln.seq<=17332`, for the 37 XXLS modified sources use condition `glorxxls.seq<=18080`.

Change	Affected sources
north33	
modified	200230 200402 200979 201037 201614
	201974 203883 204685 205950 207940
	210533 215846 219208 221849 221945
	225540 225830 226934 226951 227217
new	200123 200392 200836 200837 200848
	202822 203513 203674 203850 203856
	203860 203864 203877 203881 204275
	204276 204279 204290 205825 207428
	207932 207954 210426 210430 212739
	212860 213575 215150 215183 215635
	217136 217521 218707 220123 222029
	222040 222075 222245 223297 223528
	226895
lost	203920 replaced by 203864
	203936 replaced by 203877
	208014 replaced by 207954
	213502 replaced by 213575
	216108 replaced by 226895
	216257 replaced by 216328 by 226938
	219069 replaced by 200123
south33	
modified	200238 200248 201684 201696 201698
	201709 202965 204799 206598 207263
	208079 208284 210035 210083 210160
	210949 212124 213698 214255 214796
	217375 217711 217809 217810 217927
	218202 219297 220410 221177 221379
	221572 222239 223416 223442 223707
	223734 223883
new	207177 208311 210033 210034 210036
	210038 210043 210045 210058 210062
	210071 210076 210085 210087 210143
	210145 210161 210171 211691 214236
	214237 214798 216709 216829 216837
	217126 218097 219168 219296 221377
	221569 221641 226276 227139
lost	205251 replaced by 216709
	206110 replaced by 216829
	207262 replaced by 207177
	208879 replaced by 210058
	210048 replaced by 210076
	210073 replaced by 210145
	221296 replaced by 221377
	221517 replaced by 221641
	211385 now spurious

4.3. Data products

4.3.1. X-ray products

Note that the Xamin FITS catalogues are associated as data product to tables north33 and south33. The original association would become partially incorrect, as it points to the 3.3 FITS files only, while the `v332=1` sources contain values taken from the 3.3.2 results. As shown in the example in Fig. 1 we now supply access *also* to the 3.3.2 FITS files as an *additional data product* clearly labelled as such. The 3.3.2 FITS files are shown only if the pointing contains at least one source updated because of XAMIN 3.3.2, i.e. flagged `v332=1`. It is left to the user to look up autonomously the relevant data in the appropriate FITS file according to the `id` and `v332`.

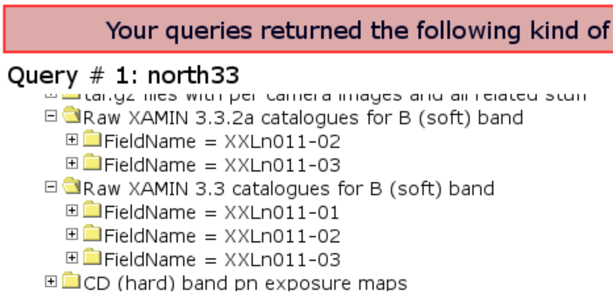

Fig. 1. Example of a query returning XAMIN FITS catalogues for version 3.3 for all pointings and for version 3.3.2 only if the pointing contains at least one source with `v332=1`

4.3.2. Thumbnail images

Thumbnail images accessed as data products of `XXLNOPT` and `XXLSOPT` or via the tool described in Chiappetti (2014b) were retrieved for the original XAMIN 3.3 list. They have now been updated with new ones for new `v332=1` sources, and have been re-centred for `v332=1` sources which changed position.

Contextually with the download of the updated thumbnails, some naming convention issues had to be fixed. In most cases this was done aligning mismatched new files to the previous conventions about surveys, subsurveys and magnitude bands (Table 4 and section 4.1.2 of Report XVI). The only exception is represented by the optical (BVRiz) survey in the SXDS area, previously called UDS in Report XVI and now called XUDS for *all* files.

4.4. Multiwave catalogues

4.4.1. "Lucio's" version

With the repointing of the GCTs for `XXLNOPT` and `XXLSOPT` (see 3.8.3), the latter two multiwave catalogues are up to date consistent with the rest of the database aligned to XAMIN 3.3.2.

This can be of limited interest since `XXLNOPT` and `XXLSOPT` were intended just as a working example, they do not provide any ranking of counterpart sets (although somebody might have undertaken it privately), and they used as member tables only those released in the database at end September 2013. However the update was due for consistency and completeness.

In `XXLNOPT` the presence of 135 new counterpart sets is the trivial effect of new 3.3.2 objecs or of the displacement of the X-ray position for `v332=1` sources. However there are other differences. One can go a bit deeper generating an `unid`, an *unique identifier* ofeach counterpart set, which concatenates the `seq` in all member tables. 53221 entries (96% of the total) preserve the same `unid`, i.e. the same counterparts are confirmed.

The truly lost entries (see 4.2) are 27 for 7 X-ray sources, but there are further 2277 entries whose `unid` has disappeared, for 611 X-ray sources. Of these only 23 entries for 14 X-ray sources are due to `v332=1` X-ray position displacement. These cases have been inspected only by sample, and are found to be due most likely to changes in the non-X-ray tables (updates) done after Sep 2013 (i.e. the generation of the original `XXLNOPT`). For instance `marseillespec` was updatedwith new entries, so we have nominally different `unids` where a `marseillespec` member was added to the pre-existing, unchanged members. It cannot be excluded however that in some cases the order of execution of the scripts has caused different choices (for instance in the association of CFHTLS `w1t7` counterparts to the other tables).

The complement of the lost entries are 2439 new `unids` for 660 X-ray sources. Of these 41 are new `v332=1` objects, and the rest are possibly `unids` changed in one or a few members,which go along with the equivalent "lost" ones. As a curiosity, if we ignore the `marseillespec` table, which had a lot of updates (insertions) after Sep 2013, the mismatches reduce to 1685 lost and 1898 new.

For `XXLSOPT` we have 80 trivial new entries; 23636 (84%) confirmed counterpart sets wiht same `unid`; 4635 lost entries for 2952 X-ray sources (of which 9 truly lost wiht 23 counterpart sets, and 30 which changed position because of 3.3.2, with 43 counterpart sets), the rest are `unid` changes complemented by 4642 new entries for 2955 X-rau sources, of which 34 are new in 3.3.2 and 34 affected by position displacement.

Considered the provisional nature of the catalogues we do not investigate any further.

4.4.2. "Sotiria's" version

With the repointing of the GCTs for `TESTXON` and `TESTXOS` (see 3.8.3) and the storage of the new thumbnails (see4.3.2), the latter *pseudo*catalogues are aligned with XAMIN 3.3.2. They provide ranking of counterparts `2XLSS`-style, and use member tables as described in Report XIV.

Note however that it is not intended to release these catalogues in the database (but only a future better version based on the final version of "Sotiria's tables");

a comma-separated (CSV) dump has been provided to Sotiria Fotopoulou (as done in March for the original version). They can however be accessed via the tool described in Report XVI.

Note that this update *formally invalidates* the numeric values of the statistics reported in Report XIV. It is not however felt worthwhile to update them in detail, considered the unofficial nature of both releases, and the limited changes. We may instead say just what follows.

What is immediately apparent if one makes a breakdown of the counterpart sets by rank, is that `TESTXON` now has many more rank 0 cases w.r.t. rank 1 (5215 vs 1662), while it used to have a more balanced distribution (3889 rank 0 vs 2980 rank 1) if we confine to the most plausible cases (flagged `autorank=0`). This does not occur for `TESTXOS` (it has 2333 and 1588 rank 0 or 1, and it had respectively 2202 and 1757). For some reason the repeat of the procedure for the north was "more favourable".

An effective comparison between the old and new version can be restricted to the "best" counterparts, i.e. those ranked 0 or 1 (by construction there is only *one* with such ranks per X-ray source). The `unid` in this case is constructed only from "Sotiria's" member tables (which are those used for ranking, see Report XIV).

For `TESTXON` we easily locate the 7 lost and 41 new objects. We are left with 14127 X-ray sources of which 14020 (99.3% !) have the same `unid`, i.e. confirm exactly the same preferred counterparts (although only 12562 are ranked identically, namely 1420 are promoted from rank 1 to rank 0, and 38 demoted from rank 0 to rank 1). The remaining 107 cases have partially or totally different `unid`s. Only 8 of them are `v332=1` sources: 3 of them have a position change over 6″with the choice of an altogether different counterpart set, the rest changed only between 0.3 and 2.5″. In general the changes in `unid` is the addition or swap of one counterpart while all the rest of the `unid` remains unchanged, or is the choice as rank 0/1 of a former rank 2 (secondary) or rank -1 (rejected) counterpart set. If one considers *all ranks* (primaries, secondaries and rejected), 92% of the counterpart sets are identical.

An alternate rough comparison can be done on the extracted CSV files, which measure the different lines (due to a change in counterparts in the counterpart set, insertions or deletions): 2721 different entries (3.8%), 250 lost entries and 96 new entries.

For `TESTXOS` once easily located the 9 lost and 34 new entries, one is left with 11854 cases of which 11268 (95%) have the same `unid` (although only 11286 are ranked identically; 233 are promoted from rank 0 to rank 1, and 108 demoted from 0 to 1). The remaining cases are 226 (of which just 10 are `v332=1`, 3 with distance changes above 6″, and the rest between 0.7 and 4.4″), otherwise similar to what reported for the north. If one considers *all ranks* (primaries, secondaries and rejected), 94% of the counterpart sets are identical.

The rough comparison on the extracted CSV files reports 4557 different entries (7.7%), 207 lost entries and 57 new entries.

Considered the provisional nature of these catalogues we do not investigate any further.

Acknowledgements. I acknowledge the supply of updated data products by Sotiria Fotopoulou.

References

Chiappetti, L., 2013a, Comparing XXL and 2XLSSd catalogues, A preliminary report, XMM-LSS Internal Report N. 11-Mi (Report XI)

Chiappetti, L., 2013b, The XXLN and XXLS catalogues, Preliminary release for internal use, XMM-LSS Internal Report N. 12-Mi (Report XII)

Chiappetti, L., 2014a, Steps towards identification of XXL X-ray sources aka the Geneva exercise 0.0, XMM-LSS Internal Report N. 14-Mi (Report XIV)

Chiappetti, L., 2014b, Accessing XXL thumbnails and catalogues, A java standalone tool, XMM-LSS Internal Report N. 16-Mi (Report XVI)

Faccioli, L., Pierre, M., Clerc, N., Pacaud, F., 2014a, Latest improvements in pipeline development and definition of the P1 class, Version 1.3, January 30, 2014

Faccioli, L., Pierre, M., Clerc, N., Pacaud, F., 2014b, Recovering point sources when the point source fit fails, Version 1.1, March 6, 2014 (on wiki).