XMM-LSS INTERNAL REPORT n°20-Mi
November 22, 2021

The latest XXL X-ray processing

Xamin 4.3 inthe Milan DB

L.Chiappetti[1]

associated, INAF, IASF Milano, via Corti 12, I-20133 Milano, Italy

Abstract. We present the 4.3 Xamin release, give basic information about the pipeline, the ingestion procedure in the Milan DB, the database tables and a comparison with the previous official (3.3) and interim (4.2) release. The new release was used preliminarily for the Loto exercise in summer 2021, and is now being released superseding the interim 4.2, which should be retired.

Key words: XXL; database tables

1. Introduction

This document describes the actions done in the Milan DB for the ingestion (and subsequent actions) of the results of the latest, hopefully final, XXL pipeline (known as Xamin 4.3 or `XaminF18`).

For details on the inner working of the pipeline refer to Faccioli et al. 2018, hereafter Paper XXIV. For the changes un current version (4.3.4) refer to the presentation by Sunayana Bhargava at the latest consortium meeting.

This report is arranged as follows: the next subsection of this introduction (1.1) gives an historical excursus; section 2 provides summary information about Xamin and the data used (in 3.3, 4.2 and 4.3); section 3 describes the steps of the procedure; section 4 describes the 4.3 database tables; and section 5 provides a comparison between 3.3, 4.2 and 4.3 results.

Note that *the PDF version of this report provides clickable links* to web pages like wiki pages and articles.

1.1. Historical excursus

I recall here the events between June 2021 and the date of print, with particular regard for what concerns database tables.

- An historical excursus of *earlier versions* of `XaminF18` pipeline (Xamin 4.0 to 4.2) is given in section 1.1 of Chiappetti (2019), hereafter Report XIX.
- 4.3 basic differences w.r.t. 4.2 are reported on the wiki in the 2021 meeting presentation by Sunayana Bhargava.
- A check with the 4.3 pipeline applied to a few individual pointings was done in May 2021, to conclude we could apply the 4.2-computed astrometric offsets (see Report XIX) to build the tiles for 4.3.
- The results of Xamin 4.3 for the *soft X-ray band* were provided by Sunayana Bhargava in June 2021, and ingested in the Milan database with an ”emergency procedure” in order to provide a reference dataset of extended sources for the Loto exercise done during the summer in view of the optical observing campaign in September. They were presented in summary form at the virtual consortium meeting in July 2021.
- During this stage the recipes for AC sources were tuned.
- The results of Xamin 4.3 for the *hard X-ray band* became ready later, and have now been ingested in the Milan DB, and merged with the soft band data, superseding the emergency ingestion.

2. Preliminary information

The *main* differences of the `XaminF18` w.r.t. the Xamin 3.3.2 used for the publication of the `3XLSS` catalogue (Chiappetti et al. 2018, hereafter Paper XXVII) were described in Report XIX, and are recalled briefly here. The differences between 4.2 (Report XIX) and 4.3 (present version) are *hilighted*. *In the remainder where version 4.x is mentioned, this means the item is equally applicable to 4.2 and 4.3 versions. 4.3-specific items are explicitly indicated.*

- More pointings were used as input to the tiles (see 2.1). The same pointings were used for 4.2 and 4.3.
- X-ray observation data are *no longer* processed by individual *pointings* but by 1-square-degree tiles (actually $68' \times 68'$ allowing some overlap at the edges). *Tuning in wavelet filtering, masking/segmentation in Sextractor, fitting of the position only for the extended source case, and some tuning to the AC and C1/C2 recipes were implemented by Sunayana Bhargava for 4.3.*

– As a corollary the handling of the overlap was different between 3.3 and 4.2, and is the same for 4.3.
– As a further corollary in 4.x there can be more than 999 per XAMIN FITS catalogue files (identifiers require four digits not three).
– In addition to the two customary "pointlike" and "extended" fits, two more fits are done for each detection, a "double" (DBL) fit considering two point sources, and an "extended+pointlike" (EPN or AC) fit considering an extended source with a central AGN.
– *The recipe for AC has been modified and tuned for 4.3. Details are given in section 3.1. An AC flag can assume values 1 or 2 as specified there. A source can be classified as "full AC" (`class='A'`, capital) if the AC flag is non-zero and is a C1 or C2, while it is a "small AC" (`class='a'`, lowercase) if only the AC flag is set but is not C1 nor C2.*
– *Also in 4.3 the C1 and C2 recipes have been tuned with respect to Paper XXIV, now the condition on the extension parameter `corerad` aka EXT is `corerad` > 3″, instead of previous 5″.* The rest of the recipes is unchanged.
– *Finally in 4.3 the likelihood threshold to designate as non-spurious the non-extended sources above it has been raised from 15 to 20.*
– Also, while XAMIN 3.x worked inside 13′ in each pointing (which allowed to consider only the MOS+pn fits in ingestion), with 4.x is now possible to have MOS-only or pn-only detections (although few).
– Astrometric correction (see 3.4) is no longer a post-processing (Milan) task but is anticipated in the earlier stages of the pipeline (Saclay task). The same corrections are used for 4.2 and 4.3.

2.1. Pointings and tiles

As said repeatedly in many places, in XAMIN 4.x XMM pointings are combined in 1-deg tiles (actually $68' \times 68'$). The layout of the 37 XXL-N and 31 XXL-S tiles is reported in the St Jacut presentation on the wiki. Here we summarize only the geometric arrangement for XXL-N:

	01	02	03	04	05			
06	07	08	09	10	11	12	13	14
15	16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31	32
						33	34	35
							36	37

And for XXL-S:

	01	02	03	04	05
06	07	08	09	10	11
12	13	14	15	16	17
18	19	20	21	22	23
24	25	26	27	28	29
30	31				

Information on the pointings as used up to DR2, whose list is published as Appendix B in Paper XXVII, were (and remain) available in the Milan DB in table `xxlpointings`.

The new list of pointings used as input to XAMIN 4.x are contained in the equivalent table `xxl4pointings`, which will be released with the rest, and accessible also correlated with the merged X-ray tables. The following info is copied verbatim from Report XIX.

The new list contains 363 XXL-N and 272 XXL-S pointings. They cannot be immediately compared with those in the old list (294 and 328) because of the following reasons. Our own internal naming convention (the tables supply also the correspondence with ESA convention's obsId) uses names of the form XXLx*mmm*-*ppc* (e.g. XXLn000-01a) where: the single letter x is either "n" or "s" for XXL-N and XXL-S: the three-digit code *mmm* is the pseudo-mosaic identifiers, i.e. either is the same for all pointings using the newer mosaic-pointing mode (i.e. XXLn115-10b) or for all older pointings of a former program (like XXLn000-01a); the individual pointing identifier *pp* designates a specific pointing direction request within a given mosaic or program; the repeat code *c* indicates different observations of the same pointing direction, because the previous ones had an insufficient exposure for various reasons (this is a letter a,b,c,... in chronological order of the repeat). Note that on a first instance the convention was not used systematically in DR2: some "first repeats" and most "single repeats" (directions pointed only once) had a blank repeat code. Now, since 4.2, all those cases have systematically a repeat code of "a" (so old XXLn001-01 is now XXLn001-01a, etc.) : the same pointing has a different name. On a second instance the old list included 81 pseudo-pointings with a repeat code of "z" (e.g. XXLs998-18z) which weren't actual pointings but the combination of 2 or more repeats which were very similar in pointing diretion and roll and could be combined in a way prototypal of the current tiling coaddition. These pseudo-pointings are obsoleted by the new tiling procedure and therefore do not exist any more. So the true number of DR2 *real* pointings was 278 XXL-N and 263 XXL-S.

These clearly shows that the new list contains a number of *newer pointings*. They can be due to newer observations of our own in AO13 after the DR2 observations; to newer archival (i.e. not our own) observations (like XXLn999-10a, 11a, and 12a: the code `mmm`=999 groups all archival observations); or to newer programs (the Brandt observations with codes *mmm*=995, 996 and 997). In addition few bad pointings might have been considered not eligible for the tiles and are now not listed. The quoted pointing tables contain the nominal pointing and roll coordinates (FITS columns suffixed `PNT`).

I note that 21 pointings were not used in any 4.x processing i.e. not included in any tile: 7 of them have bad quality (no exposure); 3 have also very limited, MOS-only exposure; but the remainder

have significant, often rather high (up to even 80 ks) exposures. They are all archival (*mmm*=999) observations: 8 XXL-N (02, 04, 07, 09-14) and 2 XXL-S (01 and 03). I hope there was a valid reason to exclude them.

There is also a working table (not released) called `xxl4centres` *which contains the coordinates of the pointing centres from the FITS keywords* `RACENO` *and* `DECCENO`. It was used during the *previous* ingestion of the XAMIN 4.2 individual pointing data to compute the off-axis angle (these data were used for the computation of the astrometric corrections). *This procedure was not repeated for 4.3*

Moreover the *FITS pointing files* (new since XAMIN 4.2) are supplied *one per tile* and have several records for each *id* in the tile. Usually these go in groups of 5. Each group contains 5 "fits" (the PNT, EXT, EPN and DBL fit plus the background BCK), or exceptionally less. There is one group per camera combination (MOSPN, MOS, PN) so normally there will be 15 entries (unless some combination was not processed by XAMIN), and there is one such set of 15 or less for each pointing contributing to the tiles.

I have ingested those files *for 4.3* in four *not released* tables, called `northb`*n*`pointings`*43* and `southb`*n*`pointings`*43*,where band n=2 or 3 for soft and hard bands. The tables contain the `id`, the camera combination and fit and the pointing *plus the tile number* (since otherwise the id won't be unique).

The tables are sort of bulky and complex to be used in conjunction with the X-ray data tables (one has to correlate on (tile,id) as well as on the same camera combination and preferred fit used by a given detection in its preferred classification). So I chose to do such operation on the fly myself and add *also for 4.3* to the X-ray merged tables `north43` and `south43` two columns called `nptgb` and `nptgcd` which contain the number of pointings used for the specific source in the soft and hard band.

Usually the two numbers are equal (unless one is zero because no detection occurred in one band). However there are a few cases (194 in XXL-N and 116 in XXL-S) in which the two bands used a different number of pointings (the difference is usually 1, anyhow never more than 2 pointings). The impression is that it is physiological and not a band merging artifact, as the inter-band distance `maxdist` is usually low. However 26 XXL-N and 19 XXL-S have `maxdist`$> 6''$, which could be suspect.

Currently the columns `nptgb` *and* `nptgcd` *are not advertised in the menus, but their name can be inserted manually in the "advanced queries".* `nptgb` for extended sources is present in the XXLDB used for the Loto.

In addition the pointing files were used once to build a correlation table between `xxl4pointing` *and the band-merged tables* `north43` *or* `south43`.

If anybody feels to need to use detailed information on which specific pointings are used by which source, please contact me and we will arrange a way.

3. Processing steps

The input supplied by Saclay is, as before, in form of *FITS catalogues*, one per tile and per band (in addition there are additional *FITS pointing files* which contain the info, for each detection, of the pointings contributing to its position). Information about usage of pointing files in the database is provided in section 2.1.

The format of the 4.X catalogues is similar to the old XAMIN 3.3.2 with a substantially higher number of columns (112 instead of 87), since there are families of parameters for the DBL and EPN fit, mimicking the layout for PNT and EXT.

As before a number of columns is actually unused or uninteresting (this includes e.g. all the error columns which are undefined *and vestigial EPN_ML_EXT which I ingested but should be ignored*).

The columns concerning the off-axis angle are no longer present (since the concept of off-axis angle w.r.t. the pointing centre is not applicable to tiles).

3.1. Ingestion

The ingestion script `bandingest43.csh` has been mimicked on the earlier XAMIN 3.3.2 and 4.2 versions.

It handles tile names instead of pointing names, it deals with the new columns as specified in section 4.1, and implements the new classification scheme, based on the following 4.3 recipes (only P1 recipe is unchanged w.r.t. previous versions). *So far check that* `detlik_ext` *is not NaN is not implemented.*

- C1 recipe: `c1c2=1` where
  `corerad>3 and extlike>33 and detlik_ext>32`
- C2 recipe: `c1c2=2` where
  `corerad>3 and extlike between 15 and 33`
- P1 recipe: `p1=1` where
  `detlik_pnt>30 and (corerad<3 or extlike=0)`
- class `P` (pointlike) as default unless one of the following is set
- class `E` (extended) initially where C1 or C2
- *AC flag set to* `ac=1` *where*
  `epn_corerad >=5 and epnlike_ext>=10 and epnlike_pnt>=20 and epn_ratio>=0.2`
- *AC flag set to* `ac=2` *where not already 1 and*
  `epn_corerad >=5 and epnlike_pnt>=20 and epn_ratio>=0.2`
- *class* `a` *("small" EPN aka AC) set where*
  `ac>0`
- class `A` ("full" EPN aka AC) reset where
  `class='E' and ac>0`
- class `D` (double) reset where
  `class='E' and dbllike>extlike`
- *this means class D and A are assigned only to sources previously classified as extended, but class a instead to non-C1 non-C2 sources*

– non-spurious sources: spurious=0 are by definition those classified E, A or D plus the P ones with detlik_pnt>*20*

The ingestion script since 4.2 onwards has provisions to handle *detections in single cameras*. This does not require any change to the SQL code, but acts on the temporary ASCII input file generated from the FITS XAMIN catalogue. While before such input file took only the MOS+pn entries, now it is the concatenation of MOS+pn entries for the cases where they do exist, plus MOS-only or pn-only ones where the MOS+pn one does not exist, handled by three separated awk scripts guided by a "preanalysed" detection list.

3.2. Band merging

The band merging script bandmerge43.csh has been mimicked on the earlier XAMIN 3.3.2 and 4.2 versions.

Besides obvious changes because of table layout (see section 4.2), it generalizes the old logics to the new 4.x classification scheme. *In addition for 4.3 it implements the "spurious threshold" at a likelihood of 20, and handles both AC classes A and a.*

The logics is that one merges a soft and a hard source where their *reference positions* are within 10″. The reference position is the one corresponding to the classification *in the band*.

So if a source is classified EP the soft extended position is compared with the hard pointlike position, if it is classified PA the soft pointlike position is compared with the hard EPN position, if it is PP or AA positions from the same classification are used, etc.

Of course sources detected in a single band (E-,-P etc.) are also brought along.

Once a source is merged, it can be *reclassified*. This is tagged by flag reclass. A final value of reclass=0 means the original classification is preserved, while reclass=2 means the source was reclassified. Reclassification means that the classification in the soft band prevails.

Reclassification is rather *rare* (51 XXL-N and 48 XXL-S cases).

So a source classified EP will be considered as if it were EE (extended), one classified PA or PD as if it were PP (pointlike), an AD, AE or AP as if it were AA (EPN). This means the merged reference values (detection likelihood per band, coordinates, counts, rates[1]) are taken from each band based on the reclassification.

So for instance an EP was considered as such because the extended soft position was within 10″ from the pointlike hard position. But the positions stored in the band-merged tables will be both from the extended fit. So at the end the inter-band distance maxdist *might* exceed 10″. This is flagged by flag suspect=1.

This is actually very rare (7 cases in XXL-N and 7 in XXL-S).

Note that the extent parameter(s) coreradb|cd are taken from the EPN fit for class A (and a), and from the EXT fit otherwise.

Note that the "ultimate" (now "intrinsically" astrometrically corrected) coordinates ra,decl are taken from the (reclassified) coordinates in the "best band" bandid (=2 or 3 for soft and hard respectively). This is the band with the best (highest) detection likelihood. A corollary of this is that if both soft and hard detection likelihoods are undefined, the source remains with bandid=0 and has no coordinates assigned. *Note that even if the detection likelihood is undefined, a coordinate in the band (hidden columns _rab,_decb or _racd,_deccd) may have a value, but so far this is not copied into ra,decl.*

All the above (already there in 4.2) is procedurally the same as for XAMIN 3.3.2. The number of bandid=0 cases is 174 in XXL-N and 54 in XXL-S (slightly more than in 3.3).

The following auxiliary definitions may instead be different in the newer scripts.

The extended flag is *now* set to 1 for sources (re)classified E or A (and even a) or D.

The merged spurious flag is set to 0 (non-spurious) for cases flagged extended=1 or for those having the (reclassified) detection likelihood above 20 in at least one band. The combination is the same as in previous scripts, but since the threshold is now higher, statistics can be affected.

The merged c1c2 is similarly set to 2 or 1 if at least one of the bands is respectively C2 or C1 and the source has extended=1.

The *ambiguities in band merging* are dealt with in the same way as in the previous version of the scripts. If two soft sources match the same hard source, or two hard sources match the same soft source, they are initially flagged suspect=2 (later the flag is transformed in a pointer to the associated source in the post-processing stage). *XXL-N gives 8 couples with a soft-to-same-hard or vv. ambiguity, and none in XXL-S, definitely less than for* north33 *or* south33*.*

3.3. Post-processing

Traditional post-processing did occur in this order for XAMIN 3.3:

– The "divorce and repoint partner" procedure for suspect=2 merging ambiguities is now performed essentially as for 3.3.
– The computation of position errors based on a tabulation of rates and off-axis angle *Since this has not been*

[1] Note that summed counts and exposure-averaged rates present in the band-merged tables are historical *(and for this reason are no longer advertised)* and are not the ones used in the public catalogues which are *going to be* computed independently from individual band values

verified by simulations and the off-axis is not defined for tiles, this step will ***NOT*** *be performed*.

- the computation of fluxes from rates using "standard" CFs. is now performed essentially as for 3.3.
- The computation of errors on fluxes is now performed essentially as for 3.3. These standard tasks are briefly recalled below.
- The astrometric correction *is no longer a post-processing task as it is anticipated in the event file preparation, care of Saclay, before* XAMIN. Some elements are however given in 3.4.
- The tile overlap removal is discussed in 3.5
- A final step provides a pointer to the `3XLSS` sources (see 3.6).

3.3.1. Divorce-and-repoint

The small cases of band merging ambiguities is handled essentially as for 3.3 (the only difference is that now merged id's are now 8-digit strings).

The ambiguity occurs when two merged sources come from the same soft source (their id's will be `ssssaaaa` and `ssssbbbb`), or conversely from the same hard source (their id's will be `xxxxhhhh` and `yyyyhhhh`). In this case if one of the cases has an inter-band distance `maxdist` above 6″ and the other is below 6″, it is assumed that the largest distance case is an independent soft or hard source (and therefore it is *divorced*, it's id is reset to `0000bbbb` or `yyyy0000`, and the `suspect` flag is set to a negative value, the `seq` of the divorced partner changed of sign). If both inter-band distances are above or below 6″, the ambiguity is considered unresolved and the `suspect` flag is positive, the `seq` of the partner. The partner remains anyhow a two-band detection (de facto they are all `PP` pointlike).

In XXL-N of 16 cases (8 couples), 6 are the divorce of a soft source, none of a hard source, and two (sources 608928 and 608929, id's 01040090 and 01290090, and sources 617151 and 617152, id's 03700306 and 03980306) remain unresolved.

In XXL-S there were no cases at all.

3.3.2. Fluxes and errors

The flux computation script has been adapted from the previous flux and position error computation one, scrapping out the latter part.

It uses the same conversion factors used in Paper XXVII and earlier, computed *for a* $\Gamma = 1.7$ *powerlaw with* $N_H = 2.6 \times 10^{20} cm^{-2}$. As before, the flux is undefined for C1 sources, and is computed from *pointlike rates in all other cases* irrespective of classification !

The only 4.x difference is that for MOS-only or pn-only sources the flux is computed from single camera rates, while in the other case it is, as before, the mean of the MOS and pn fluxes.

The flux error computation script is also essentially the same used for Paper XXVII, propagating the errors on reconstructed number of gross counts, with adequate provisions for the cases of MOS-only or pn-only sources.

More details are documented in a note on the wiki.

3.4. Astrometric correction

The astrometric correction, in form of a per-pointing rigid shift, is now applied before XAMIN when merging pointing event files into tile event files.

The astrometric offsets were computed beforehand for each 4.2 pointing *(this was done only at the time of 4.2 processing and what follows is taken verbatim from Report XIX)* in a way *analogue* to what done in the past using the SAS task `eposcorr` (with no rotation), but with the following differences:

- XAMIN catalogues produced using the 4.2 pipeline *on single pointings and in the soft band only* were used (database tables `north42ptgb` and `south42ptgb`).
- The reference X-ray list for each pointing consisted therefore of non-spurious pointlike `soft` sources (all, as baseline, and those with off-axis angle less than 10′ as auxiliary set).
- Since `eposcorr` requires an estimate of the X-ray position errors, they were computed using the offaxis-rate tabulation used in Paper XXVII and earlier.
- The reference optical list (unlike Paper XXVII, which had used the "official" CFHTLS or BCS photometry) uses the latest available photometry extraction by Sotiria Fotopoulou (2015, database tables `SFNcfht`, `SFSbcs` and `SFSdecam`, see Chiappetti 2016, hereafter Report XVIII), i.e. CFHT (including our own ABC fields) in XXL-N and BCS or DECam in XXL-S. For each X-ray source in case of multiple tentative counterparts it takes the one with the best chance probability computed using the density tabulations described in Report XVIII, taking sources brighter than magnitude 25 (bands y or i for CFHTLS, band g for ABC, band i for BCS and DECam).

The results of the computation of the astrometric coefficient has been supplied to Christian Garrel as ASCII files, and is also available for documentation as web pages for XXL-N (all data), XXL-N (10′ offaxis only), XXL-S (all data), XXL-S (10′ offaxis only).

Note that for some (poor) pointings `eposcorr` may not run at all if there are no X-ray sources meeting the criteria in the pointing or no optical reference objects are found (these are missing altogether from the web pages, "non-made"). Also the result of `eposcorr` might be less reliable in the case the number of X-ray sources (and therefore optical references) in the pointing is small, which usually results in computing offsets but not assigning an error to them ("zero-error").

My advice was (similarly to what done in the past) to do no correction at all for such pointings. Also restricting the X-ray reference lists below 10′ increases the number of such cases (from 8 zero-error and 6 non-made to 12 and 8 in XXL-N, from 9 zero-error to 14 zero-error and 1 non-made in XXL-S) therefore I advised to prefer those based on the "all source" reference list.

I have no idea whether some of those "very bad" pointings were to be excluded from the tiles already for other reasons, so I do not know which pointings were uncorrected or ignored.

For pointing XXLn999-01a (the Mira Ceti field where no CFHT observations were made) the astrometric offsets were computed using the USNO catalogue.

For 4.3 a sample check on a few pointings was made, and the conclusion was to use for the tiles the same per-pointing astrometric offsets computed for 4.2.

3.5. Overlap removal

In the past (i.e. up to 3.3.2) overlap removal was done on band-merged data after astrometric correction (and removed detection in multiple overlapping pointings), keeping with priority the detection from the "best field" (and, if two pointings were equally good - or equally bad, preferring the one with the smallest off-axis angle). The removal did not actually delete anything from the "physical tables" (like `north33`) but simply created *catalogue views* (like `XXLN`[2]) which made visible only the overlap-surviving sources. Contextually also all spurious detection were "removed" from the catalogue views.

Until a catalogue view is create, overlap removal is limited to the provision of a flag for deletion to the detections which should be removed. Hence, in the merged tables `north43` *and* `south43` *only, one can limit to the overlap-surviving sources using the clause* `deleted=0`. The `deleted` flag assumes two other possible values (-1 for the spurious detections and 1 for the non-spurious redundant detections which should be removed).

I considered as potential overlaps the cases where two detections observed in different tiles (avoiding counting couples (a,b) and (b,a) twice) have their *reference* coordinate closer than 10″. Coordinates are natively astrometrically corrected. Spurious detections and detections with `bandid=0` (actually included within the spurious ones) are ignored from the very start.

Considering a criterion for overlap removal, I considered this. The overlap between two adjacent tiles is a strip, therefore one can expect that the detection in one tile could be close to the edge of the strip (and therefore be problematic), but the 10″-associated detection in the other tile would be towards the inside of the tile.

[2] The Paper XXVII published catalogue `3XLSS` is the *union* of the northern and southern portions `XXLN` and `XXLS`

So I considered both the nominal distance to the tile centre, but also the individual RA and DEC offsets. If the offset in DEC is about 1 deg, and the offset in RA is small (tiles at same RA) one should prefer the source with smallest DEC offset. If the offset in RA is about 1 deg, and the offset in DEC is small (tiles at same DEC) one should prefer the source with smallest RA offset. Finally if the two tiles are not at same RA or same DEC, consider the distance to the tile centre and take the smallest (this caters for areas at the corner of tiles where a source may be detected 3 or 4 times).

Since rejection occurs in couples, one has to make sure that in each multiplet one and only one detection survives. Multiplets can be couples (a,b,0,0), triples (a,b,c,0) or quadruples (a,b,c,d). Of course one shall get rid of redundant multiplets (for instance if there is a quadruple (a,b,c,d) the couples (a,b,0,0), (b,c,0,0), (b,d,0,0) and (c,d,0,0), and the triples (a,b,c,0) and (b,c,d,0) shall be ignored). Then one applies the tile-centre distance criterion to each couple and builds a 4-char flag *xyzt* : each char in the flag is initialized to `-` (undeleted) or `.` (if the corresponding position is unused, i.e. last after a couple or triple), set to `D` when deleted and to `x` when a deletion is propagated. So a couple can be either `-D..` or `D-..` (first or second survives), a triple `-DD.`, `-Dx.` etc. and the flag shall contain one and only one `-`: cases like `-D-x` or `x--D` or `DD..` are anomalous.

All the above is taken verbatim from Report XIX as the same overlap handling procedure is used. Only the few numbers below are specific of 4.3.

Despite all criteria used I found no way to get an anomaly-free automatic solution, so I listed the cases with anomalous flags (1 in XXL-N and 24 in XXL-S) and manually checked and arbitrated them. Actually I had one "preparation script", one adhoc fixup script (incorporating the manual arbitration) and one "deletion flag application script", so the procedure at the end was still rather automatic.

In `north43` the overlap-surviving detections are 12222, the to-be-deleted overlaps are 2810, and the spurious cases to be ignored 18843. In `south43` the corresponding figures are 8608, 1914 and 15700. *Note the relatively higher number of spurious sources, due to the likelihood threshold increase from 15 to 20.*

A further systematic check was made for 4.3. All what follows in this section is new. If one compares the flux of the kept source with the flux of the associated deleted source(s), one finds usually a ratio of about 1. However there are cases where the kept/deleted ratio is lower, clustering around a value of 0.5.

The fluxes have often large error bars, however there is a small fraction (227 in the N and 211 in the S) where the ratio is lower than 0.7 in one or the other energy band (or both). These figures reduce to 169 and 115 with a significance above 2σ, 113 and 60 at 3σ and 64 and 22 above 5σ.

There is no obvious correlation with the detection likelihood for the deleted sources with an anomalous flux (higher, about double, than the kept one). On the other hand note that spurious sources are not considered at all in the overlap removal procedure (otherwise said if a non-spurious source would "overlap" a spurious one, it will be considered as single, not overlapping).

It is better instead to look at the RA and DEC offsets. For this purpose the RA offsets shall be corrected for projection effects ($cos\delta$ is about 0.5 in the XXL-S area). This way the offsets are always in the range 0-34′ (note that if a source is close to a tile *side* one of the offsets will be in the range 30-34′, and the other is unconstrained; if the source is close to a tile *corner* both offsets are large).

Now, as it is apparent from Fig. 17[3] if one considers the largest between RA and DEC offsets (i.e. the one closest to a tile side) one clearly sees that the sources with a kept/deleted flux ratio around 0.5 are all at offsets larger than 33′ (i.e. at the very tile edge). On the contrary the offset of kept source is virtually never above 33′ (2 cases in XXL-N and 1 in XXL-S …in these cases one could imagine the fluxes are *equally wrong*). We are instead justified to consider the (lower) flux of the kept sources as the correct ones.

3.6. Pointers to XAMIN *3.3 sources*

It is rather important to keep track of the association of new 4.3 detections to previous 3.3 detections. To be *general enough* I am providing a "preferred association" in term of a column `3xlsspointer` present in `north43` and `south43`. When non-zero it contains the `seq` of the preferred source in `north33` or `south33`.

What follows is taken verbatim from Report XIX *except for the replacement of the references of 4.2 to 4.3* since the same procedure is used.

This is not the *most general* solution, which may involve the use of *correlation tables* (see 4.4) and allow for multiple associations. Note that in fact there are issues like spurious detections, multiple detections in tile overlaps and multiple detections in overlapping pointings. Plus the fact `north33` and `south33` have raw and astrometrically corrected coordinates.

I started (for both XXL-N and XXL-S) from a couple of provisional correlation tables (i.e. the "corrected" one listing the associations with a distance below 10″ between `north43` coordinates and `north33` corrected coordinates, and the "uncorrected" one listing the associations between `north43` coordinates and `north33` raw coordinates). One can have cases like:

- no 3.3 counterpart either using corrected or uncorrected
- one 3.3 counterpart in one case and none in the other
- one 3.3 counterpart in both cases and it is the same, or it is different (this in practice does not occur)
- more than one 3.3 counterpart in one case and none in the other
- more than one in one case and one in the other (in practice always the same)
- more than one in both (equal or different)

The critical cases are those with more than one counterpart. In principle one should prefer "the closest" but …closest in corrected or uncorrected distance ? And also, what about if the closest was removed by the 3.3 overlap removal procedure ?

For each 4.3 source with more than one 3.3 counterpart, I flagged three conditions: (a) the couple is closest in corrected distance; (b) the couple is closest in uncorrected distance; (c) the 3.3 source is in the `3XLSS` catalogue (i.e. non-spurious overlap-surviving). I then considered a priority: those with all three conditions set are preferred, then the `3XLSS` closest-corrected, those closest-uncorrected, and those just in `3XLSS`, then those not in `3XLSS` with the two closest distance, and finally those with just the closest corrected distance: this way all 4.3 source have *at least one* preferred counterpart.

I said "*at least one*" because it is possible to have more than one with the same priority. In this case a tie-break is made, usually taking the couple with the smallest distance (when this is not possible because the distances are the same, because the two counterparts are "suspect ambiguous" and share the same distance, the tie-break just takes arbitrarily the one with the snallest `seq`).

In XXL-N, of 33875 sources, 15883 have no `north33` counterpart (of these 14034 are spurious), and 17992 have a preferred counterpart (of these only 4809 are spurious): those where the counterpart is in `3XLSS` are 12591 non-spurious and 2415 spurious.

In XXL-S, of 26222 sources, 11065 have no `south33` counterpart (of these 10481 are spurious), and 15157 have a preferred counterpart (of these only 5219 are spurious): those where the counterpart is in `3XLSS` are 9679 non-spurious and 2820 spurious.

A corollary of all above is that if you query a merged table for a given value of `3xlsspointer` and you are returned more than one entry, one of them will have `deleted=0` and the other will be flagged for deletion as they are due to tile overlap.

3.7. The "emergency procedure"

This entire section describes the "emergency procedure" ran in early summer 2021 to enable the later Loto exercise. None of the interim tables (which are preserved) will be released via the Milan DB. Make reference only to the final results within the Lyon XXLDB.

The emergency procedure included the following steps:

[3] Figures for this section are reported at the end of the document, in order not to change the numbering of all subsequent figures w.r.t. Report XIX (for ease of comparison).

- ingesting the soft band data: these are the *existing and released* tables `north43b` and `north43cd` or `south43b` and `south43cd`,
- a fake band merge and post-processing was done, but the resulting tables (though stored offline) were *overwritten* by current `north43` and `south43`.
- an *old reference dataset* was made including all objects in pre-existing XXLDB (`lyon` in Milan DB) and all nominally extended objects in 3.3
- a *4.3 reference dataset* was made including: (a) all extended objects in 4.3; (b) all 4.3 objects with a *3xlsspointer* in the old reference; (c) 4.3 objects within 30″ from XXLDB objects keeping all tagged extended and anyhow the closest one; (d) residual `3xlsspointer` and objects in the old reference not yet located.
- classify the objects based on flags proposed by Marguerite Pierre and technical flags dealing with the various overlap conditions (3.3, XXLDB and 4.3!), and assigning a `megakeep` flag (0=ignore; 1=good combinations; 2=recovered XXLDB primary sources; 3=recovered 4.3 extended sources so that no one of the latter two categories is lost).
- pass an interface file to Saclay with all the interesting parameters

3.7.1. Technicalities

This is mainly for my own use if I had to go back and dig out old files (or scripts).

The *old reference dataset* is generated by scripts `make`*xxx*`th4333.sql` and is stored in tables *xxx*`th33cluref` (with *xxx*=`nor|sou`, here and in the remainder). These tables have the `seq` in `lyon` and *xxx*`th33`, plus some interesting parameter from the latter (columns with names starting in `(O)` for "old").

The *4.3 reference dataset* is generated by scripts `make`*xxx*`th33cluref.sql` and is stored in tables *xxx*`emergency`. These tables have the `seq` in `lyon`, *xxx*`th33`, and in 4.3 plus some interesting parameter from the latter (columns with names starting in `(N)` for "new"). Note that `NBseq` is *current* *xxx*`th43b.seq`, while `NXseq` points in the *removed fake merged table*.

The *interface files* (or better the tables from which comma-separated files are generated) are created via commands in `README.createtestinterface` and then edited via `flagif.sql`. They work on a temporarily-named file to be renamed at the end. The files exist in two incarnations *xxx*`thtestif` and `bis`*xxx*`thtestif` (the latter has one more column and a different flagging and is the latest form sent to Tatyana Sadibekova at Saclay). These files contain the columns requested for the Loto, some additional technical flags, and the pointers `Lseq` into XXLDB `lyon`, `OXseq` into *xxx*`th33` and `NXseq` into the *removed fake merged table*.

The objects used for the Loto exercise were chosen among those with `megakeep>0` according to some combinations of `selectflag`. This flag is in the form of 5 characters which can be `-` (unset) or a number 1-5 in the position set (more positions can be set, e.g. `1-34-` or `-2--5`). Position 2 can be replaced by an `X` for the few sources marked "c3" in XXLDB (XXL-N only). The flags are intended for

- 1 4.3 C1/C2 sources new not in XXLDB
- 2 XXLDB sources lost (not present or not extended) in 4.3
- 3 C1/C2 changed in 4.3 w.r.t 3.3 or XXLDB
- 4 AC also C1/C2 (class *A*)
- 5 AC not C1/C2 (class *a*)

4. The updated database

The names and layout of the database tables for `XaminF18` are similar to the ones for XAMIN 3.3.2, but for the obvious addition and removal of new or disappeared columns.

The old numeric `field` number has been discontinued (there is only a mnemonic `FieldName` in form `XXL`x`Tile-`nn, with x=`n|s` and nn=1-37 or 1-31 respectively), as well as the `badfield` flag. Also all astrometric-correction related columns have been discontinued since the coordinates are natively corrected.

The source sequence number `seq` *starts at 600000 for XXL-N and 700000 for XXL-S*, so that the two are natively distinct between them, and with respect to earlier pipeline version.

Note that, while such ***`seq`*** *from* `north43` *and* `south43` *will be used as* ***`Xseq`*** *in the catalogue views, it has* ***nothing to do*** *with the* ***`NXseq`*** *used in the interface files provided in the "emergency phase" for XXLDB and Loto. The latter refers to an earlier version of a fake band merged table with just the soft sources, and does not correspond to any pointer in the current tables. Use exclusively the combination of Fieldname and* ***`id`*** *in the soft-band tables* `north43b` *and* `south43b` *to match the Loto* ***`Ncid`****.*

4.1. Individual band tables

Individual band tables (respectively soft and hard) are called `north43b` and `north43cd` for XXL-N and `south43b` and `south43cd` for XXL-S.

As customary they include all (MOS+PN) entries in the FITS catalogue, *plus now* the MOS-only or pn-only cases (they are rather few, 88+376 in `north43b`, 119+292 in `north43cd`, 36+427 in `south43b` and 46+337 in `south43cd`; they can be identified because the pn or MOS exposures are zero).

The mapping of database column names to the FITS column names (which in turn are not always the names in Paper XXIV) is sort-of hybrid. In most cases for historical reasons I maintained the names used for earlier versions of XAMIN, which used a suffix `_ext` or `_pnt` to which I now added `_dbl` and `_epn`, as specified in Table 1, but for

Table 1. Mapping of database column table names to FITS column names for "suffixed" families. The number *nn* in front of FITS column names are the `TTYP`*nn* keyword values. Taken verbatim from Report XIX.

database name	*suf*=`_pnt`	*suf*=`_ext`	*suf*=`_dbl`	*suf*=`_epn`
`detlik_`*suf*	13 PNT_DET_ML	14 EXT_DET_ML	16 DBL_DET_ML	15 EPN_DET_ML
`x_`*suf*	18 PNT_X_IMA	41 EXT_X_IMA	93 DBL_X_IMA	67 EPN_X_IMA
`y_`*suf*	20 PNT_Y_IMA	43 EXT_Y_IMA	95 DBL_Y_IMA	69 EPN_Y_IMA
`ra_`*suf*	22 PNT_RA	45 EXT_RA	97 DBL_RA	71 EPN_RA
`dec_`*suf*	23 PNT_DEC	46 EXT_DEC	98 DBL_DEC	72 EPN_DEC
`ratem_`*suf*	25 PNT_RATE_MOS	48 EXT_RATE_MOS	100 DBL_RATE_MOS	74 EPN_RATE_MOS
`ratep_`*suf*	27 PNT_RATE_PN	50 EXT_RATE_PN	102 DBL_RATE_PN	76 EPN_RATE_PN
`countm_`*suf*	29 PNT_SCTS_MOS	52 EXT_SCTS_MOS	104 DBL_SCTS_MOS	78 EPN_SCTS_MOS
`countp_`*suf*	31 PNT_SCTS_PN	54 EXT_SCTS_PN	106 DBL_SCTS_PN	80 EPN_SCTS_PN
`bkgm_`*suf*	33 PNT_BG_MAP_MOS	56 EXT_BG_MAP_MOS	108 DBL_BG_MAP_MOS	82 EPN_BG_MAP_MOS
`bkgp_`*suf*	34 PNT_BG_MAP_PN	57 EXT_BG_MAP_PN	109 DBL_BG_MAP_PN	83 EPN_BG_MAP_PN
`pixdev_`*suf*	35 PNT_PIX_DEV	58 EXT_PIX_DEV	110 DBL_PIX_DEV	84 EPN_PIX_DEV

altogether new columns I used straightly the FITS names (although in lower case). All nomenclature in this section inclusive of Table 1 is taken verbatim from Report XIX.

The following column names are historical or mimicked on historical (the number between the database name and the FITS name is *nn* from the FITS `TTYP`*nn* keyword values).

- cutrad 37 PNT_CUTRAD[4]
- extlike 38 EXT_LIKE
- corerad 39 EXT
- dbblike 87 DBL_LIKE
- separation 88 DBL_SEP
- theta 92 DBL_THETA
- epn_corerad 63 EPN_EXT

epnlike_ext 61 EPN_LIKE_EXT and epnlike_pnt 62 EPN_LIKE_PNT are a hybrid case.

The following maintain the FITS name in lower case dbl_ratio 90, epn_ratio 65, and epn_ml_ext 17 *(vestigial)*.

As usual NaN values are replaced with -1.

4.2. Band merged tables

Band merged tables are called `north43` and `south43`.

They include all detections in all tiles (i.e. also the multiple ones in the 8′ overlap strips between tiles), including spurious ones (i.e. any detection likelihood).

They have the same layout as `north33` and `south33`, but for the removal of obvious columns and for the fact the merged `id` is now 8- instead of 6-character long (the `id` is the concatenation of soft and hard id. Since now each tile can have more than 999 detections, the individual id can occupy 4 digits, and since leading zeros are used, one has cases like `12350129` in tile XXLnTile-18 (for the merging of soft id 1235 with hard id 129) or `01060055` in tile XXLnTile-01 for soft id 106 merged with hard id 55.

[4] there are 4 values of cutrad for the 4 fits but they are all identical

The positions and fluxes in the merged table are those corresponding to the *new* classification of the source (i.e. `E`,`P`,`A`, ***`a`***, `D`).

In addition `north43` and `south43` have two new columns, `deleted` (see 3.5) and `3xlsspointer` (see 3.6).

As customary in the past, two clone tables `north43dup` and `south43dup` are provided in the case one wants to correlate one merged table with itself (e.g. to know how many sources exist in proximity of a given source). Since the current database interface does not allow to enter a table name twice in a query, this trick (e.g. correlate `north43` with `north43dup`) using a "clone view" allows it.

4.3. X-ray catalogues

Catalogues refer to source lists where spurious or multiple detections have been removed.

They are released with the names `43XXLN`, `43XXLS`, plus the ancillary band ones `43XXLNB`, `43XXLNCD`, `43XXLSB` and `43XXLSCD`.

They have the same format as the published `3XLSS` catalogue (see Paper XXVII (Chiappetti, 2018) for all details on column names etc.) or the interim 4.2 catalogues released afer Report XIX, which are retired contextually to the release of these 4.3 ones.

They inherit the correlations from their main physical member (`north43` or `south43`).

4.4. Correlation tables

4.4.1. Internal 4.3 correlations

Apart from using the catalogues, which however offer only a customary selection of columns, the only way to get the complete per-band information of sources in the merged tables is to use the correlation table labelled "*on our own XAmin identifier*". This is offered when ticking a band merged table (e.g. `north43`) and *one* of its band tables (e.g. `north43b`). *Therefore this works one band at a time.*

A 30″ correlation table between a merged table and its clone (e.g. north43 and north43dup) allows to speed up queries for sources "in proximity" of another source.

Ad-hoc correlation tables exist between xxl4pointing and the band-merged tables north43 or south43, generated as described in 2.1, which can be used to query the pointings involved in a given detected source.

4.4.2. 4.3 to 3.3 correlations

Concerning the correlation between 4.3 (north43 or south43) and 3.3 (north33 and south33) I am providing two "merged" correlation tables for each N and S. One is labelled "*new-to-old 10 arcsec corr or uncorr coordinates*", and the other one is labelled "*old-to new (reverse) ...*". They are "merged" from the temporary ones used in 3.6 i.e. consider the cases where either the corrected or uncorrected distance is below 10″. The "direct" one has at least one entry for each 4.3 source, and may return a null 3.3 entry if no 3.3 counterpart exist, otherwise it will return *all potential* 3.3 counterparts (not just the ones preferred by 3xlsspointer). The "reverse" correlation table starts from the 3.3 side, hence it has an entry for each 3.3 source, which may have zero (null), one or more 4.3 counterparts. In either cases the objects in the "second" table without counterparts in the "first" are ignored (that's why one needs separate direct and reverse tables).

4.4.3. Correlations with other tables

Correlations of north43 *and* south43 *with the existing photometric and other tables will be arranged in the near future as the need for it arises in the preliminary work towards a multiwave catalogue.*

Correlation with the XXLDB cluster "clone" table lyon *within 30″ has been generated and is available but should be regarded as coarse and provisional until the correspondence between XXLDB and 4.3 data will be studied in detail. For the time being refer to the "Comparison" sections 5.2.1.*

4.5. Data products

4.5.1. Original catalogues

The access to the original FITS catalogues (per tile and per band) *is to be arranged in the future*

4.5.2. X-ray products

X-ray images, exposure maps etc. are currently not available and presumably will be arranged in the future

5. Comparison 4.3 vs 3.3

A 4.2 to 4.3 comparison can be made indirectly comparing the numbers and plots in Report XIX with the equivalent ones in the following sections of the present report.

5.1. Basic statistics

I report below some basic statistics, usually in a three column-group format, where the first column report the item, the second group the 4.3 count, and the third group the 3.3 count. In the numeric column-groups, the first figure is for XXL-N, and the one in parenthesis for XXL-S.

First of all the total number of entries in the merged and individual band tables.

Total merged	33875	(26222)	26555	(27173)
Total soft	24876	(18544)	19352	(20067)
Total hard	15265	(12835)	13258	(12797)

Apparently 4.3 is deeper in XXL-N, but one has to check the number of *non-spurious* entries (as absolute value and percentage of total). *It is not so deep as 4.2, possibly because it is cleaner.*

Non-spurious merged	15032	(10522)	17398	(18145)
Non-spurious soft	14088	(9773)	15245	(15749)
Non-spurious hard	4484	(3565)	6848	(6759)
Non-spurious merged	44%	(40%)	66%	(67%)
Non-spurious soft	57%	(57%)	79%	(78%)
Non-spurious hard	29%	(28%)	52%	(53%)

The *number of spurious detections is definitely higher* for 4.3. *It was already higher for 4.2 (vs 3.3), and it is now even higher because the ML threshold was raised to 20.*

We consider now the various classifications first for the soft band (pnt means pointlike, ext means extended).

pnt spurious	10788	(8771)	4107	(4318)
pnt non-spurious	13793	(9579)	14873	(15306)
pnt P1	10743	(7134)	8775	(8562)
ext E	259	(162)	372	(443)
ext A (AC or EPN)	4	(3)	n/a	(n/a)
ext a (not C1/C2)	28	(23)	n/a	(n/a)
ext D (double)	4	(6)	n/a	(n/a)
ext C1	143	(87)	159	(172)
ext C2	124	(84)	213	(271)
Total ext	295	(194)	372	(443)

Then for the hard band.

pnt spurious	10781	(9270)	6410	(6038)
pnt non-spurious	4426	(3522)	6738	(6548)
pnt P1	2884	(2155)	2743	(2480)
ext E	55	(40)	110	(211)
ext A (AC or EPN)	-	(-)	n/a	(n/a)
ext a (not C1/C2)	1	(3)	n/a	(n/a)
ext D (double)	2	(-)	n/a	(n/a)
ext C1	14	(12)	29	(30)
ext C2	43	(28)	81	(181)
Total ext	58	(43)	110	(211)

And finally for the band merged case. I consider as A, a or D those which are such in at least one band.

pnt spurious	18843	(15700)	9157	(9028)
pnt non-spurious	14695	(10311)	16950	(17541)
pnt P1	11226[a]	(7481)[b]	9307[c]	(9062)[d]
ext E	299	(179)	448	(604)
ext A (AC or EPN)	4	(3)	n/a	(n/a)
ext a (not C1/C2)	28	(23)	n/a	(n/a)
ext D (double)	6	(6)	n/a	(n/a)
ext C1	152	(91)	174	(188)
ext C2	157	(98)	275	(416)
Total ext	337	(211)	448	(604)
Best band soft	24028	(17881)	18694	(19429)
Best band hard	9673	(8287)	7835	(7711)

[a] 33 P1=-1 nominally extended
[b] 24 P1=-1 nominally extended
[c] 12 P1=-1
[d] 13 P1=-1

The peculiar (seldom occurring) cases like reclessified sources, `maxdist` above 10″, `bandid=0` and merging ambiguities are counted above in sections 3.2 and 3.3.1.

We do not provide a direct comparison with the interim 4.2 data, but it can be achieved comparing the tabulations above with the corresponding ones in Report XIX.

In general one can say that 4.3 is classifying more detections as spurious, due to the increase of the likelihood threshold from 15 to 20. Also the classes D, A and a are new in XAMIN 4.x, and the number of AC sources in 4.3 w.r.t. 4.2 is much less due to the change of the recipe (see 3.1). Finally, the number of sources classified extended is somewhat less. This does not necessarily mean the source is now undetected, but it could be classified pointlike. This will be discussed later.

5.2. Lost and Found

We can use the correlation tables described in 4.4.2 to tell how many 3.3 sources are "lost" and how many 4.3 sources are "new" without a 3.3 countepart. The `bandid=0` objects are excluded by construction.

The new 4.3 sources without 3.3 counterpart are 15868 in XXL-N and 11126 in XXL-S.

The lost 3.3 sources without 4.3 counterpart are 8289 in XXL-N and 7696 in XXL=S.

The common sources are 17992 (XXL-N) and 15157 (XXL-S): the figures refers to 4.3 sources with tile overlap (the number of 3.3 sources with pointing overlap is 18426 for XXL-N and 19669 for XXL-S). In the remainder we will consider the "preferred" couples associated via `3xlsspointer` (see 3.6), i.e. 17992 XXL-N and 15157 (XXL-S).

All values refer to the band-merged tables. No specific statistics will be supplied for individual-band tables. Whenever values like counts or rates from individual-band

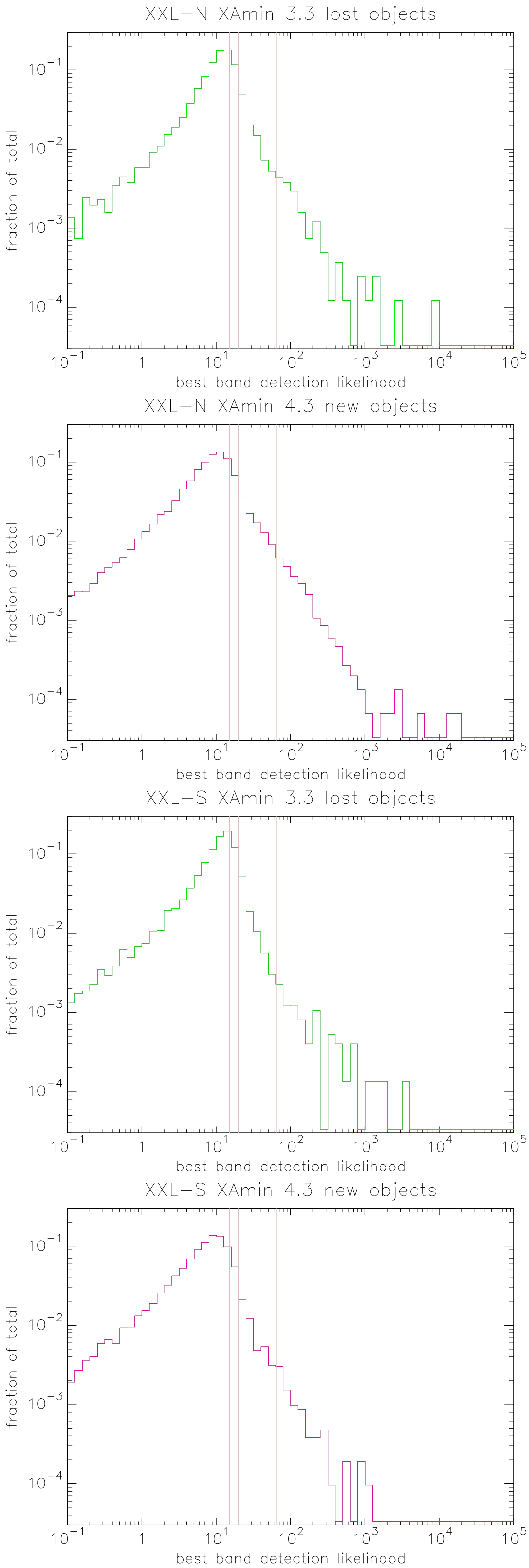

Fig. 1. Histogram of the best band detection likelihood for "lost" 3.3 detections (green histograms) or "new" 4.3 detections (red histograms), for XXL-N and XXL-S. See Fig. 4 for graphical conventions.

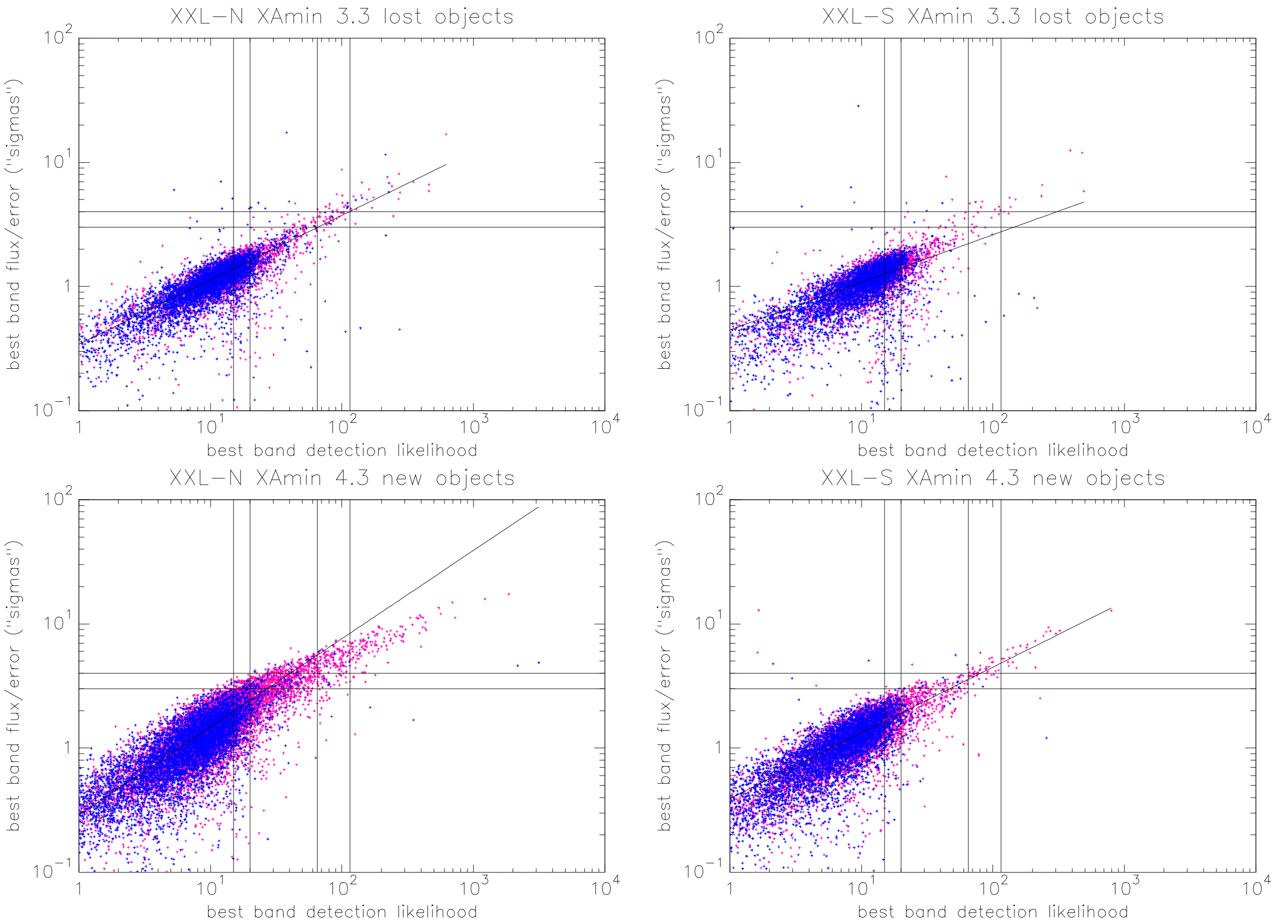

Fig. 2. Cross-calibration of the S/N vs detection likelihood (refer to caption of Fig. 5 for more details about colour coding and graphical conventions) for "lost" 3.3 detections (top row) or "new" 4.3 detections (bottom row). Left column: XXL-N; right column: XXL-S.

tables are quoted, they will be extracted using the correlation on tile and id within the tile with the merged `seq`.

It is apparent from Fig. 1 and Fig. 2 that, not surprisingly, both *the new and lost objects are concentrated among the spurious sources* and anyhow among those with relatively low detection likelihood. Note that all histograms in Fig. 1 and Fig. 4 are each normalized to the total number of elements under the curve.

It could be worth examining individually the brightest of the "lost" cases. Not unlike 4.2, there are 56 nominal C1 in the north (13 are hard-only though), and 62 C1 in the south (15 hard-only) which are lost from 3.3 to 4.3 (see next subsection for more considerations about missing extended sources).

Concerning pointlike sources, those above 4σ are 56 in XXL-N and 33 in XXL-S. Above 5σ they are 28 and 17, above 10σ 3 and 9. May be at least the latter (or their surrounding in the X-ray images) should be inspected manually. The three northern cases are 222660 (a `PP` in XXLn998-05a with a soft flux of 1.9×10^{-14} cgs; 221183 (a hard-only in XXLn115-01 with a hard flux of 2.9×10^{-12} cgs) *and 223845 (a soft-only in XXLn999-04 with a flux of* 1.3×10^{-13} *cgs), which is also the only one lost only in 4.3; all other bright lost cases were already known as lost in 4.2)*. The southern ones are too many to give details here, but 310160, 319521 and 310046 (in order of increasing significance) are soft sources in the range $5-9 \times 10^{-14}$, while 310055, 326044, 312317, 312367, 310063 and 312369 are hard-only with fluxes usually in the range $2-7 \times 10^{-13}$.

5.2.1. Extended lost and new

At face value, among the lost sources in the north there are 64 cases "nominally extended detected in the hard band only", 123 extended in the soft band, 3 extended in both bands and 6 `EP`. 120 are in XXLDB, 63 have an XLSSC number.

In the south the lost "hard nominally extended" are 132, the soft-only extended 138, the extended in both bands 3 and the `EP` 8. 97 in XXLDB of which 37 have an XLSSC number.

The number of new nominally extended sources in the north is 117. 30 are in the hard band only. In the soft band there are 80 `E-`, 1 `A-` and 2 `D-`, plus 3 `EP` and one `EE`.

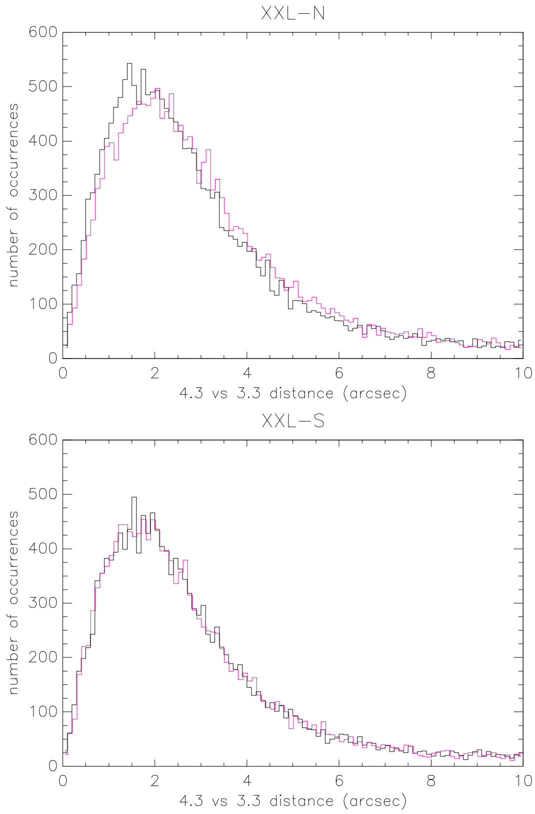

Fig. 3. Histogram of the distance between the 4.3 position and the 3.3 astrometrically corrected (black) or uncorrected (magenta) position.

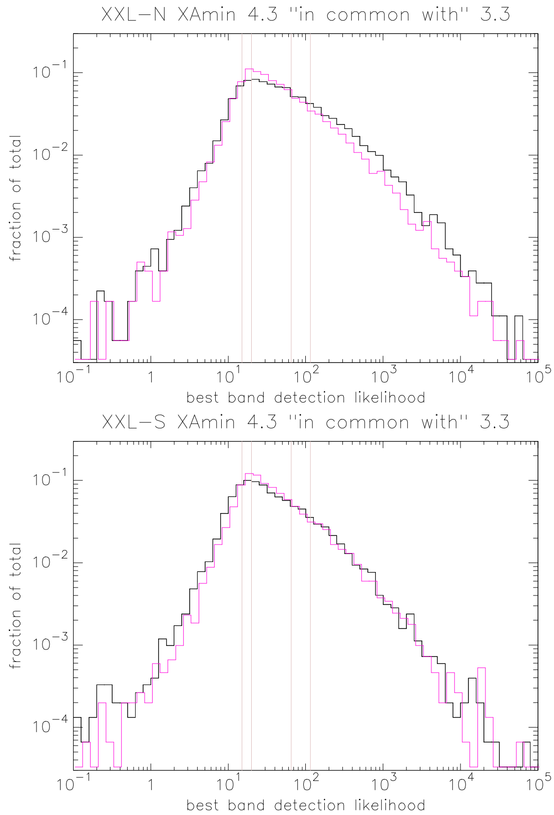

Fig. 4. Histogram of the best band detection likelihood for objects in common between XAMIN 4.3 and 3.3. The thick black curve refers to 4.3 likelihoods, the violet curve to 3.3 ones. The vertical fiducial lines are for likelihood 15, *20*, 65 and 115 (see caption of Fig. 5).

The equivalent number in the south is 48. 9 are hard-only. In the soft band one has 33 `E-`, 2 `EE`, 2 `D-` and one each of `EP` and `AP`.

One can also do a different exercise for extended sources in the Lyon XXLDB. On one hand (using the clone table `lyon`) we know the 3.3 source natively associated to each tagged cluster candidate (`Xseq`), and, when different because of the pointing overlap removal procedure, the catalogued 3.3 source associated (`truXseq`), on another hand we could correlate `north43` or `south43` with `lyon` on a larger radius (30″), and check whether the 4.3 association matches the 3.3 one using the `3xlsspointer` (see 3.6).

24 northern and 16 southern XXLDB clusters have no 4.3 counterpart within 30″: of these 10 and 7 have an XLSSC number.

199 N clusters and 177 S clusters have one 4.3 counterpart but only 141 and 133 match the `3xlsspointer` (92 and 65 with XLSSC number). 2 northern objects previously not associated now have a 4.3 counterpart. In the other cases either there is no `3xlsspointer` or it does not match the association with XXLDB.

Of the 122 N clusters and 103 S clusters with more 4.3 counterparts within 30″, 91 and 81 have one matching the `3xlsspointer` and 1 northern object previously not associated has a 4.3 counterpart.

All the above was one very coarsely repeating for 4.3 what done for 4.2, just for a comparison with Report XIX. What follows is instead based on the procedure described in 3.7 (and therefore ignores the hard band).

We consider only the cases tagged `megakeep>0` and group coarsely by `selectflag` (see 3.7.1).

- the untagged cases (`selectflag='-----'`), i.e. those which match an XXLDB bona fide candidate cluster with the same classification (E C1 or C2) are 109 in XXL-n (and 79 in XXL-S; in the remainder we will always give the XXL-S in parenthesis after the XXL-N value). 102 (70) are confirmed XLSSC clusters. 7 (9) have another status (tentative, photometric, provisional or blank) and therefore no XLSSC number yet. Note that 21 (10) cases have `3xlsspointer` not set. This means the 3.3 counterpart (from XXLDB) has a distance greater than 10″ from the 4.3 object (up to 30″).
- the objects tagged with a select flag of 1 (i.e. potentially "new") are 82 (51). A number of them are not really new sources (which would be

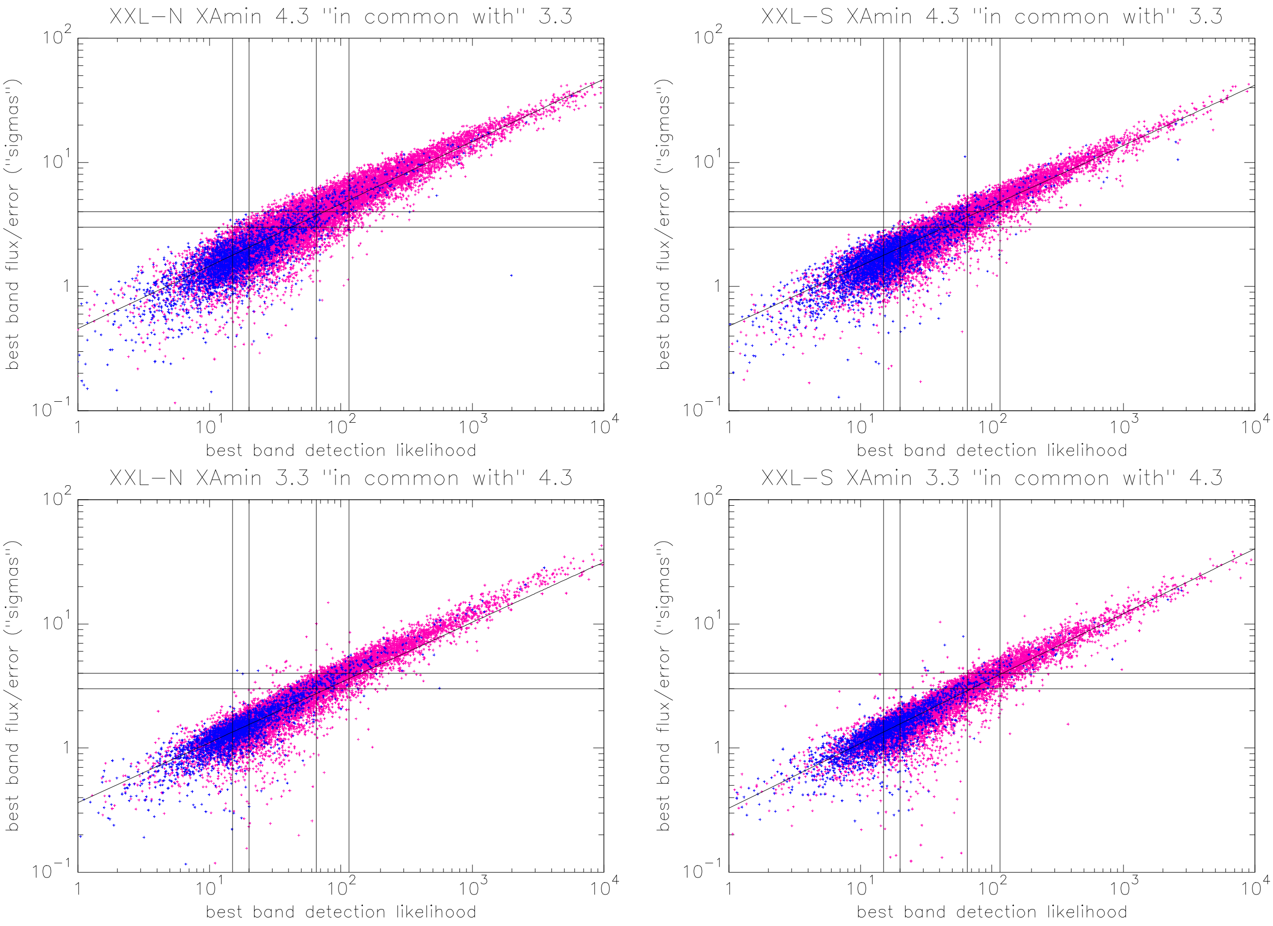

Fig. 5. Cross-calibration of the S/N (flux divided by computed flux error) vs detection likelihood in the best band for objects in common between XAMIN 4.3 and 3.3. Left column: XXL-N; right column: XXL-S. Top row: 4.3 data; bottom row: 3.3 data. In all plots red means best band is soft and blue best band is hard. The diagonal line is a linear fit in loglog space to the S/N averaged in pseudo-logarithmic bins ((the pseudo-logarithmic binning is a spacing of 1 in likelihood up to 100, then 5 up to 1000, 50 up to 10 000, and 500 above 10 000). The vertical fiducial lines are for likelihood of 15 *and 20* (below which sources *respectively were, in 3.3, and are, in 4.3,* considered spurious), 65 and 115, matching the levels of 3 and 4σ *in Paper XXVII* identified by the horizontal lines. The fit is done independently for each plot, while the fiducial lines are the same for all plots, and use the calibration established in Fig. 1 of Paper XXVII. The difference between the plots in Paper XXVII and those in the current report is that here we mix soft and hard data, and consider either one according to what is the best band for each source.

`selectflag='1----'`), but new as extended, as they may match a 3.3 pointlike object (so they have also select flag 3 set, changed classification from no C1/C2 to either C1 or C2, textttselectflag='1-3–' or even AC, `selectflag='1-34-'`). We have 39 (24) of such cases of which 2 (0) AC. One new XXL-S object is an AC (`selectflag='1--4-'`). 44 (26) are really new (no 3.3 counterpart at all).

– the objects tagged with a selectflag of 2 (i.e. nominally "lost") are 93 (97). Most of these are not really lost, but lost as "true extended". A small number have a `selectflag='-2--5'`) i.e. a 3.3 extended XXLDB candidate cluster matches an "a" class 4.3 AC. These are 4 (3). Of the pure `selectflag='-2---'` cases only 10 (10) are XXLDB candidates with no 4.3 counterpart at all (of these 7 (7) are confirmed XLSSC clusters). The rest are objects which have a 4.3 *pointlike* counterpart either within the safe 10″ range or up to 30″. Of them 62 (66) are confirmed XLSSC clusters, while the rest has a lower status.

– a smaller number of 38 XXL-N objects have a flag "X" instead of 2 (`selectflag='-X---'` or in 2 cases `selectflag='-X--5'`). These objects were classified "c3" in XXLDB, i.e. candidate clusters but pointlike in 3.3, and have also a pointlike, or exceptionally a-class, counterpart in 4.3 within the usual distances. 26 of them are confirmed XLSSC clusters.

– A selectflag of 3 indicates a change of C1/C2 classification. However this includes the change from pointlike (`c1c2=0` to C1 or C2, which were counted

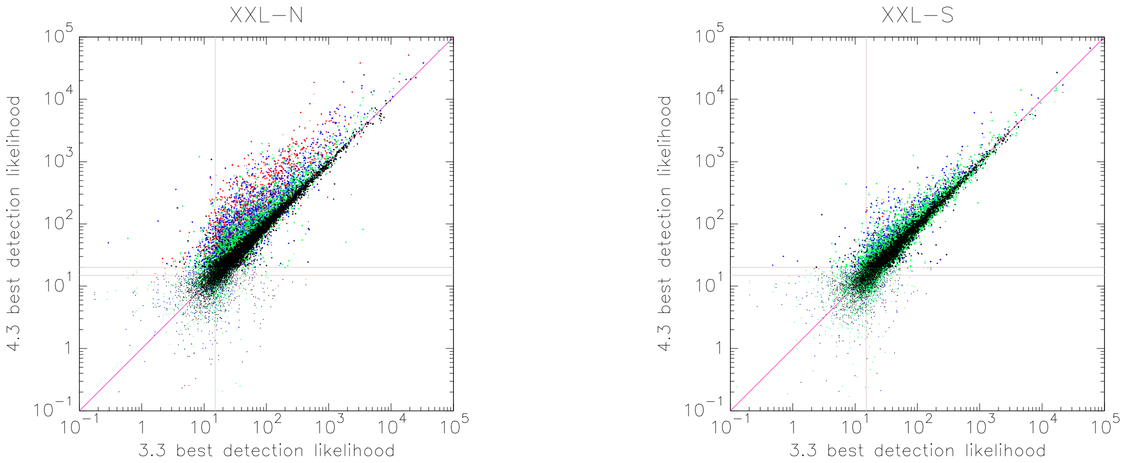

Fig. 6. Comparison of the detection likelihood in the best band for the 3.3-4.3 common pointlike sources (pointlike in both). Dots are sources not eligible for catalog (spurious or overlap-removed), crosses are those eligible (`deleted=0`). Colour codes are: black, sources detected in a single pointing; green, detected in 2 pointings; blue, detected in 3-4 pointings; red, detected in more than 4 pointings. Fiducial lines are for ML=15 *(an additional horizontal line for ML=20 is added for 4.3 data matching the new threshold)* and for equal likelihood.

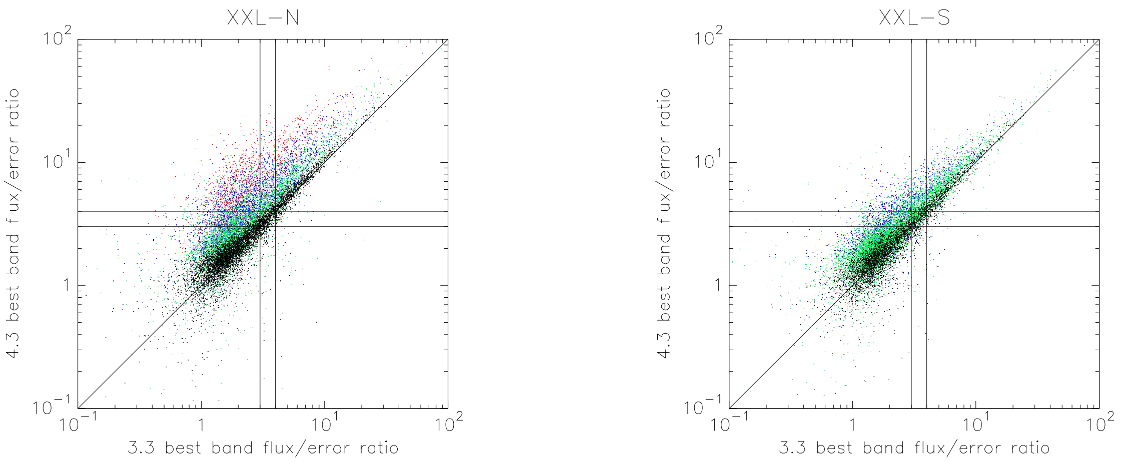

Fig. 7. Comparison of the flux/error ratio in the best band for the 3.3-4.3 common pointlike sources (pointlike in both). Colour codes as in Fig. 6. Fiducial lines are for 3 and 4σ and for equal ratio.

above within the "new as extended"). The pure plain changes are 34 (20) `selectflag='--3--'`, plus 2 (0) `selectflag='--34-'`, which are class A in 4.3. So these can be considered confirmed extended objects with minor changes. Of them 31 (14) are confirmed XLSSC clusters.

– A selectflag of 4 indicates just that the 4.3 object is an AC of class A. This (very rare) condition occurs usually in conjunction with the 1 or 3 flag, and as such has been counted above. There is a single pure `selectflag='---4-'` in XXL-S, XLSSC 516, which is now classified AC.

– A selectflag of 5 analogously indicates just that the 4.3 object is an AC of class a. This (quite rare) condition may occur in conjunction with the 2 or X flag, and as such has been counted above. A pure `selectflag='----5'` case occurs in 17 (15) cases. All these objects weren't in XXLDB, but were present in 3.3 usually as pointlike.

5.3. Common sources

In the remainder we look only at the 17992 XXL-N (15157 XXL-S) common sources associated by `3xlsspointer`. The figure in parenthesis following one not in parenthesis is always for XXL-S.

Among those 4.3 sources having a 3.3 counterpart, the number of actually independent sources (`deleted=0`) is 10677 (8106).

Among the total, 12604 (9690) are confirmed non-spurious in both 4.3 and 3.3, 2383 (2387) are confirmed spurious, 579 (248) are promoted non-spurious in 4.3 and

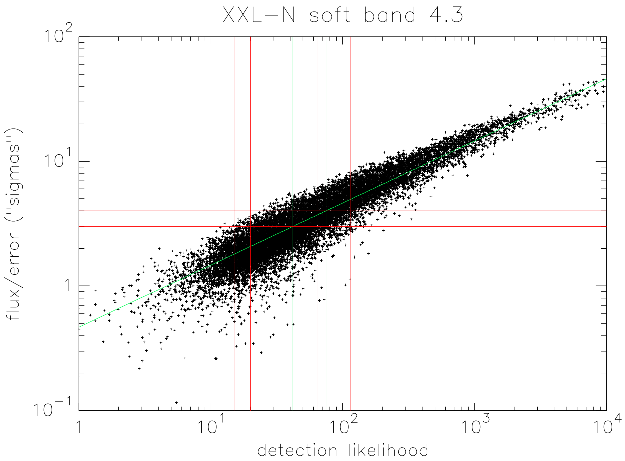

Fig. 8. Sample of cross-calibration of the S/N vs detection likelihood for a single band (soft) and dataset (4.3). Fiducial lines as in Fig. 5 except that: red lines are for the *standard* ML=65, ML=115 and 3 and 4σ levels; the green diagonal line is the fit from the *current* dataset made as in Fig. 5; the green vertical lines show the ML corresponding to 3 and 4σ based on such fit (intersection with the red horizontal lines).

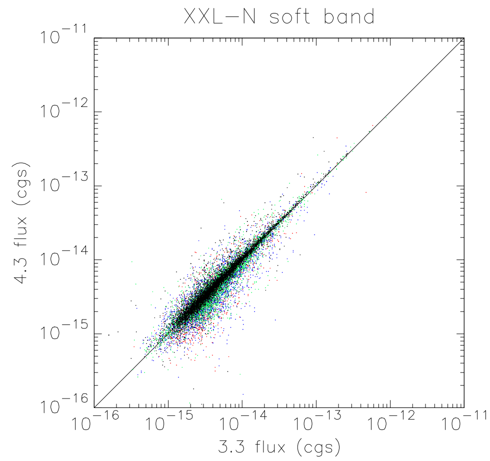

Fig. 9. Sample plot. Comparison of the flux in the soft band for the 3.3-4.3 common XXL-N pointlike sources (pointlike in both). Colour codes as in Fig. 6. Fiducial line for equal flux.

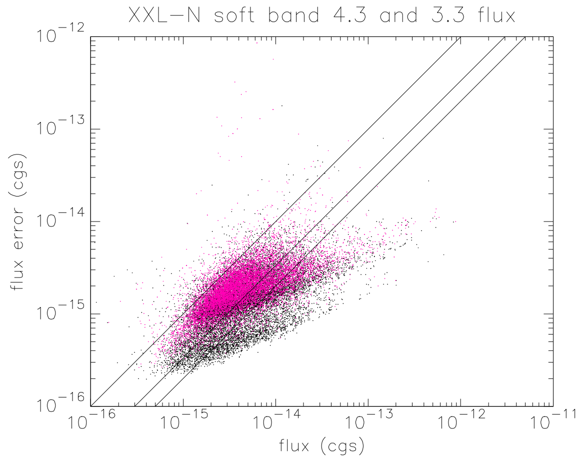

Fig. 10. Sample plot. The flux error is plotted vs the flux for the soft band for the 3.3-4.3 common XXL-N pointlike sources (pointlike in both). Black points are for 4.3 sources and magenta points for 3.3 ones. Fiducial diagonal lines are (from left to right) for significances of 1, 3 and 5σ.

2426 (2832) are demoted. *The relatively large number of demoted sources is obviously affected by the increase in the ML threshold for spuriosity to 20.*

Among the overlap-free (`deleted=0`) 10197 (7889) are confirmed non-spurious, and 480 (217) are promoted (the 4.3 spurious are to be rejected ex officio from the catalogue, `deleted=-1`).

Concerning the *generic extended flag* (which for 4.3 groups also EPN and DBL), among the total one has 17636 (14851) pointlike and 137 (115) extended which confirm their status between 4.3 and 3.3. 128 (137) 3.3 extended are pointlike in 4.3, and viceversa 91 (54) pointlike become extended.

Among those non-spurious in 4.3, 12856 (9693) maintain their pointlike status and 99 (76) 3.3 extended become pointlike (the 4.3 extended are by definition non-spurious so the other figures are identical to the total case).

Among the overlap-free, 10407 (7897) pointlike and 115 (101) extended are confirmed, 79 (45) pointlike become extended and 78 (63) extended become pointlike.

Concerning the detailed (EPDA*a*) classification, in the total one has, among the pointlike, the same classification (which in general means single- or double-band detection) for 8227 (7153) soft-only, 1918 (2148) hard-only and 3977 (3715) double-band (of which 1 (3) `PE`).

One has *compatible* classifications (e.g. `PE` vs `PP` or `P-` vs `PP` or vv.) for 1570 (1076) now detected in both bands, 1557 (495) formerly detected in both bands and now soft-only, or 122 (113) now soft-only, while 265 (151) were formerly detected only in one band and now only in the other (i.e. *not compatible*).

The above figures reduce to 1369 (885), 1448 (507), 63 (51) and 60 (14) if one limits to those non-spurious in 4.3.

For the confirmed extended, the same classification occurs just for 99 (100) cases, but this is coincidental mostly with the fact the new classification has new classes (A*a* and D). *Of 38 (15) other cases all are nominally compatible (like detection in one vs two bands, or AaD vs E) bt a single case.*

For the "new extended" a majority [39 (23)] are simply former soft pointlike now detected as soft extended, 25 (19) cases are now detected as a or A, 1 (4) cases as D, 7 (3) are miscellaneous cases now mostly detected in two bands and extended in soft, while 10 (5) are nominal extended hard sources.

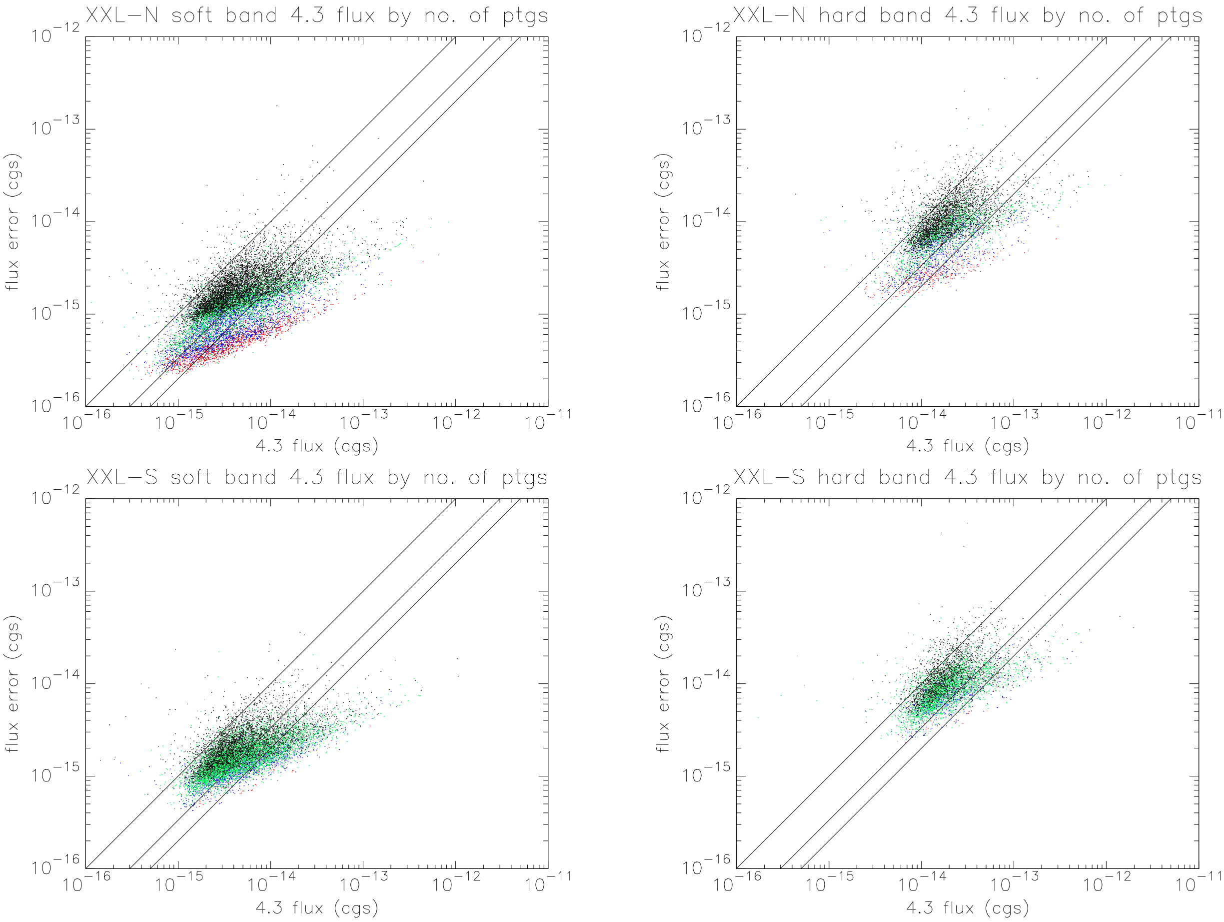

Fig. 11. The flux error is plotted vs the flux for all bands and areas, for 4.3 pointlike sources (pointlike in both 4.3 and 3.3). Colour codes as in Fig. 6 to tell the number of pointings used for each source. Fiducial diagonal lines are (from left to right) for significances of 1, 3 and 5σ.

For the "old extended now pointlike", 23 (18) were `EP` now `PP`, or 12 (6) `P-`, 11 (6) are former single-band detections now detected as `PP`, 66 (77) change from extended to pointlike in the soft band, 13 (4) are miscellaneous cases, and 15 (26) are now detected as pointlike in the hard band only.

As already said I cannot but suggest manual inspection of the "changed extended" cases.

I report in Fig. 3 the histogram of the distance between the 4.3 position and the 3.3 position. Actually the latter exists in two incarnations (astrometrically corrected and uncorrected), while the 4.3 position, as explained above, should be natively correct. To play sure I computed both distances (which however are similar within 1″ in 74% of the cases, and differ by more than 2″ in less than 3%.

As already occurred for 4.2, it is slightly surprising that there is no greater difference between corrected and uncorrected 3.3 positions (I'd expected the distance betweem 4.3 and 3.3 corrected to peak closer to zero).

Concerning the *detection likelihood*, taking the best (highest) one from the two bands, one can plot the histogram (Fig. 4) and the calibration with the flux SNR (Fig. 5), which can be compared immediately with the similar figures for "lost and new" sources (see 5.2).

One can clearly tell the difference between the "sound" common sources and the poorest lost and new sources, as well as note that the 3.3 and 4.3 behaviour is very similar (e.g. for the SNR calibration).

It is apparent that the cross-calibration between S/N and likelihood used in Paper XXVII and used as fiducial mark in the figures (see caption of Fig. 5) is not adequate in some 4.3 cases: the diagonal fitted line does not cross the 3 and 4σ level where the vertical fiducial marks do. For the entire XXL-N and XXL-S samples the levels may be something like 45 (48) and 79 (86) in likelihood, but there may be a not obvious dependency on exposure time (unfortunately the number of sources is not equispaced with exposure ranges), with the Paper XXVII levels sort-of mimicked in the 5-10 ks range.

Fig. 6 compares the detection likelihood, in the best band used above, between 3.3 and 4.3. For simplicity the comparison is limited to sources considered pointlike in both releases. The figure uses different symbols for sources considered eligible for the catalogue because surviving the

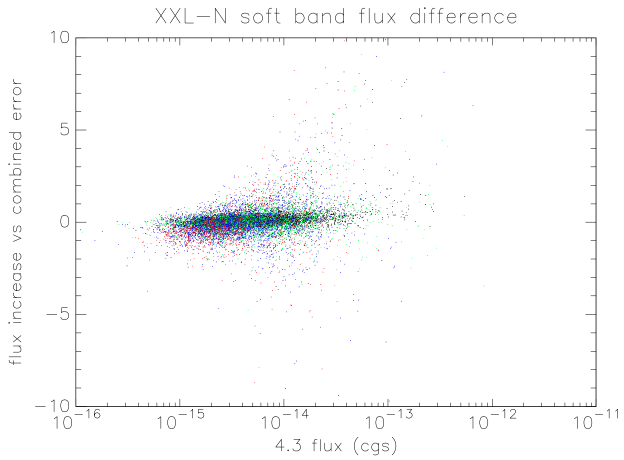

Fig. 12. Sample plot. The flux increase (or decrease, 4.3 - 3.3) divided by its error (quadratic combination of 4.3 and 3.3 flux errors) plotted vs the 4.3 flux in the soft band for XXL-N pointlike sources (pointlike in both 4.3 and 3.3). Colour codes as in Fig. 6 to tell the number of pointings used for each source.

overlap removal procedure (see 3.5) and the other (spurious or not). There is a large scatter anyhow, with *indication of a trend by which sources resulting from the coaddition of several pointings in one tile have a better (larger) likelihood* in 4.3, while those located in a single pointing have comparable likelihoods in 3.3 and 4.3. Because of the large number of sources and of the large scatter, it might be better to look at the figure at a large magnification. But even in such case points plotted later (e.g. the black single pointing after the green double pointing, and so on) may hide a number of sources plotted previously.

Note that one has that single pointing sources are 42% (33%), double are 28% (51%), triple pointings 13% (11%) etc., considering all sources. Considering catalog candidates (`deleted=0`) the percentages become 39% (32%), 28% (51%), 14% (12%) etc. (the figures in parentheses refer to XXL-S).

The improvement is even more visible in Fig. 7 where one plots the flux/error ratio in the best band. Single pointing sources give similar SNR in 3.3 and 4.3, while *sources resulting from the tiling of more pointings have a better ratio*.

I generated plots (which are not reported here in toto and are available on request) similar to those in Fig. 5 but for all detections in individual (soft and hard) bands for both 3.3 and 4.3 data. They are rather similar to the "best band" plots (see a sample in Fig. 8), except for the fact the ML level corresponding to 3 and 4σ levels may vary a bit according to case (e.g., for 4.3, 41 and 74 for XXL-N soft, or 51 and 87 for XXL-S hard), so, considered the scatter, *in the figures for fiducial marks I still report* the 3XLSS values of 65 and 115.

I omit to generate also plots per band analogue to those in Figs 6 and 7.

Similarly I report just a sample in Fig. 9 of a plot comparing the 4.3 and 3.3 fluxes in the bands. Fluxes look consistent, *almost irrespective* of the fact the 4.3 detection derives from a single pointing or from the tiling of more pointings (except perhaps for a large scatter). Adding error bars on such plots would not help, it will just make the plot less legible.

It could be more useful to plot the flux error vs the flux. Several plots can be produced, of which I report here a sample (other may be available on request). I report in Fig. 10 one case comparing 3.3 and 4.3 fluxes and errors on the same frame, and in Fig. 11 all four cases plotting 4.3 fluxes and errors according to the number of pointings used for each "tiled" source. The 4.3 errors show a large scatter than 3.3 ones and are often smaller, specially for the sources resulting from the tiling of more pointings.

As I said, the 4.3 fluxes are usually compatible with the 3.3 ones. I have computed the 4.3 minus 3.3 flux difference (increase or decrease) and found an increase in 49-52% of the cases, and a decrease in 48-51%, and a variation (increase or decrease) in absolute value greater than 100% in about 2% of the cases, greater than 50% in 5-6%, greater than 20% in 14-17%, and greater than 10% in 26-29%. The difference is usually *consistent with zero*. Fig. 12 is a sample (other figures available on request) reporting the significance of the difference vs the error (quadratic combination of flux errors)

I have also examined some data in the individual band tables, like count rates and number of counts, which are available separately for MOS and pn cameras.

The count rates behave unconspicuosly and of course similarly to the flux (Fig. 9) which derives from them. Therefore I report in Fig. 13 only two sample plots (MOS and pn) for one case, which are immediately comparable with Fig. 9, i.e. rates, as fluxes, are *very similar* between 3.3. and 4.3.

I have also generated plots of the rate/error ratio, which are available on request and of which I report here just a sample in Fig. 14, which can be compared with the rather similar Fig. 7 (left panel). I remind that the flux errors are computed by the propagation of rate errors as described in 3.3.2 and wiki page cited therein. Fig. 14 is useful just to compare MOS and pn cameras before they are weighted and combined. Otherways it shows the *usual trend* that rate/error ratio is better for sources tiled from many pointings.

Finally I report some plots about the number of counts in individual bands. The counts are available separately for MOS and pn cameras, however, unlike rates, they can be easily summed in a plain way. I report a sample of MOS and pn counts for one case in Fig. 15, and the full set of plots for summed counts in Fig. 16. In this case *the number of counts is also higher for sources tiled from many pointings*.

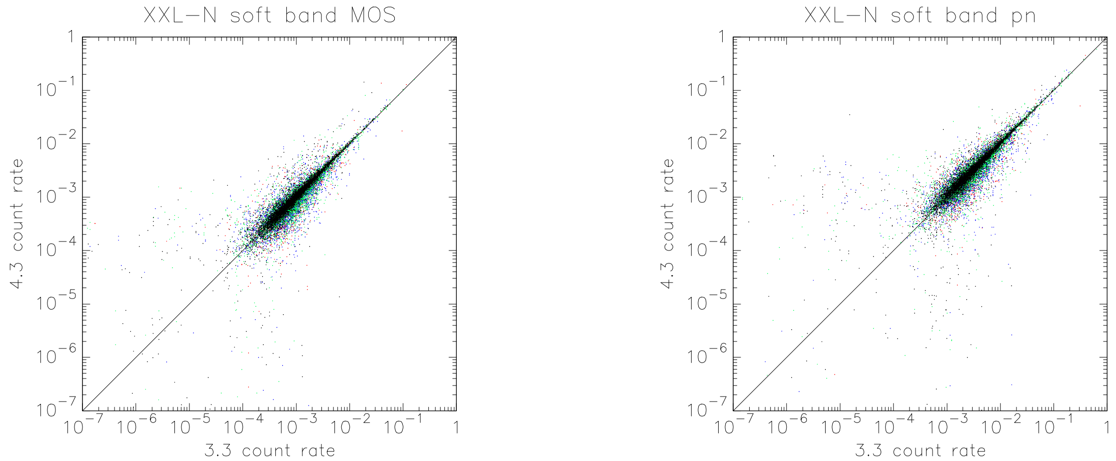

Fig. 13. Sample plot. Comparison of the count rate in the soft band for the 3.3-4.3 common XXL-N pointlike sources (pointlike in both). MOS camera on the left, pn camera on the right. Colour codes as in Fig. 6. Fiducial line for equal rate.

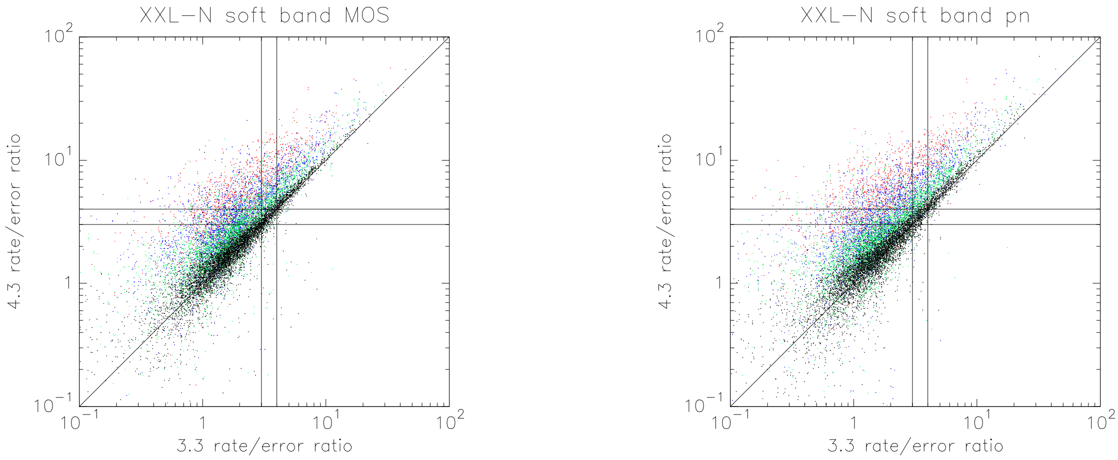

Fig. 14. Sample plot. Comparison of the rate/error ratio in the soft band for the 3.3-4.3 common pointlike sources (pointlike in both). MOS camera on the left, pn camera on the right. Colour codes as in Fig. 6. Fiducial lines are for 3 and 4σ and for equal ratio.

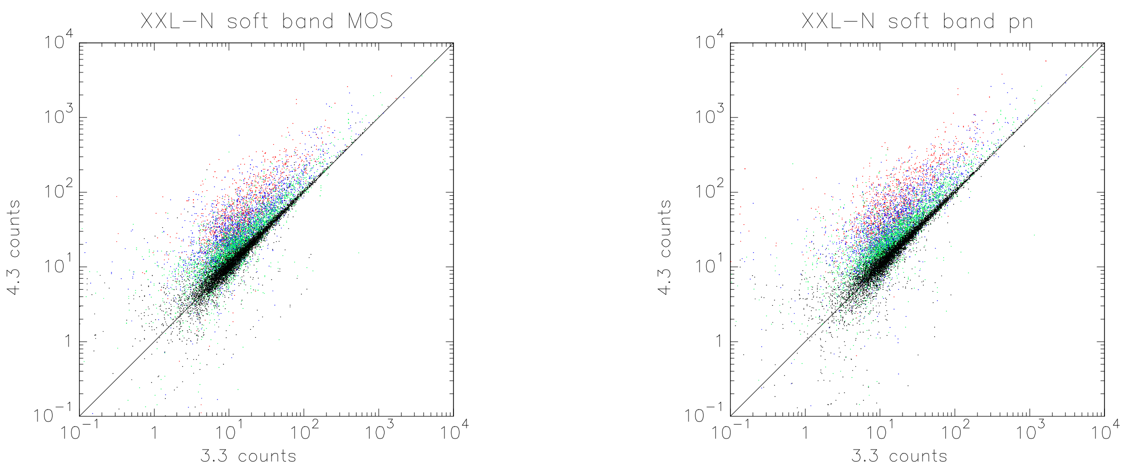

Fig. 15. Sample plot. Comparison of the number of counts in the soft band for the 3.3-4.3 common XXL-N pointlike sources (pointlike in both). MOS camera on the left, pn camera on the right. Colour codes as in Fig. 6. Fiducial line for equal counts.

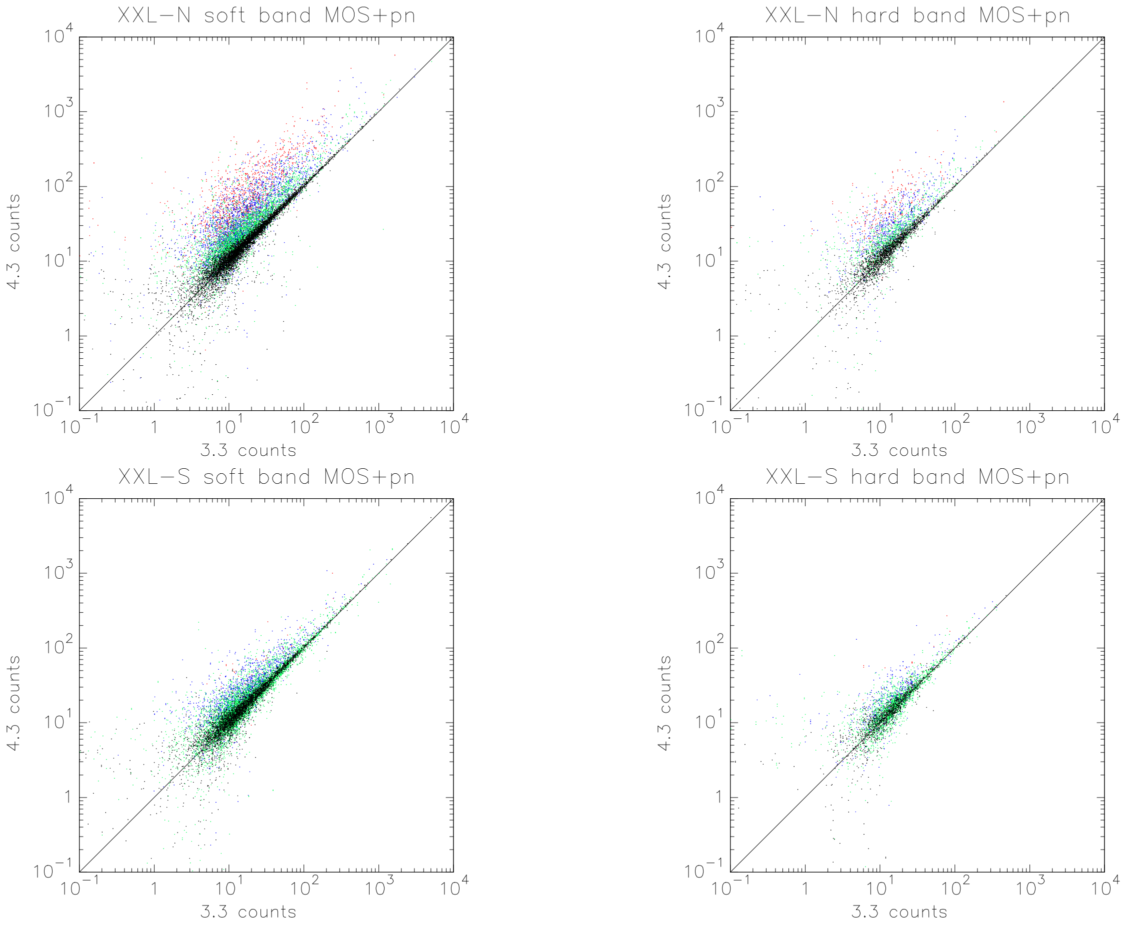

Fig. 16. Comparison of the MOS+pn summed number of counts in the soft band for the 3.3-4.3 common XXL-N pointlike sources (pointlike in both). Colour codes as in Fig. 6. Fiducial line for equal counts.

Acknowledgements. I acknowledge the work of Lorenzo Faccioli and Florian Pacaud in the development of the 4.x pipeline described in Paper XXIV and in the production of the 4.0 data; the work of Christian Garrel for the upgrade of the pipeline and the production of 4.2 data, the support of Florian Pacaud for the astrometric correction implementation, and *last but not least* *the work of Sunayana Bhargava for the upgrade of the pipeline and the production of 4.3 data*.

References

Chiappetti, L., 2016, Further steps for identification of XXL X-ray sources using the 2015 homogenized photometry. XMM-LSS Internal Report N. 18-Mi (Report XVIII)

Chiappetti, L., 2019, The almost final XXL X-ray processing. XAMIN 4.2 aka `XaminF18`. XMM-LSS Internal Report N. 19-Mi (Report XIX)

Chiappetti, L., et al. 2018, The XXL Survey XXVII. The 3XLSS point source catalogue. A&A 620, A12

Faccioli, L., et al. 2018, The XXL Survey XXIV. The final detection pipeline. A&A 620, A9

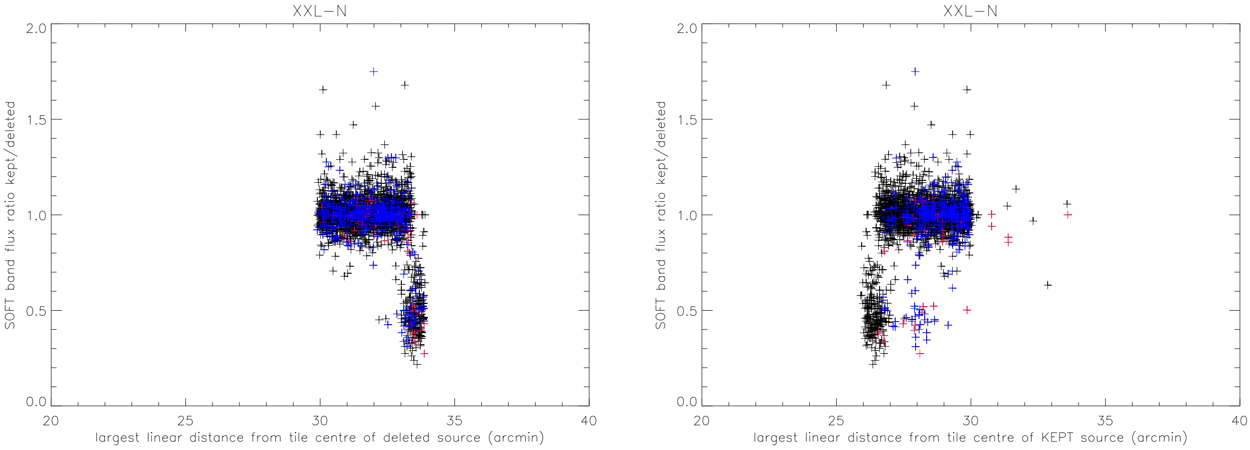

Fig. 17. Sample plot for XXL-N soft band. The ratio between the flux of the kept source and the deleted source in an overlapping couple, triple or quadruple is reported vs the largest linear distance (either along the declination or right ascension axis, the latter corrected by $cos\delta$) from the respective tile centre of the *deleted* source (left-hand panel) or of the *kept* source (right-hand panel). Colour codes are: black for couples, red for tirples, blue for quadruples.