

Further steps for identification of XXL X-ray sources using the 2015 homogenized photometry

L. Chiappetti¹

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

Abstract. I report on the ingestion in the Milan database of the photometric catalogues distributed by Sotiria Fotopoulou at the end of 2015, on their comparison with earlier catalogues, and their usage for identification of XXL sources. In particular the multiwave views released via the Milan database are also described, which allow to see the parameters of the candidate counterparts (as well as the photometric redshift and the USNO, SIMBAD and NED associations). The association of the multiwave candidates among them has been done in two ways, and for each of them tentative identification of X-ray sources has been done using two methods, chance probability and likelihood ratio, which are all compared. A quick guide on the best way to access the views is provided in section 9.

Key words: LSS; XXL

1. Introduction

The support given by the Milan DB towards the identification and characterization of XXL X-ray sources has slowly proceeded through various steps. Originally the Milan DB included *unmodified* photometric subsets in various bands drawn from public or private (consortium) sources as described in Report XII (Chiappetti, 2013). In 2014 Sotiria Fotopoulou (SF) raised the issue of the need of *homogeneous* photometric catalogues, and distributed to a small internal team a preliminary dataset. This, despite not being in a final shape, was used for the exercise described in Report XIV (Chiappetti, 2014a), available on the XXL wiki. At the Sexten consortium meeting it was agreed to have a revised dataset at end 2014, however at that time SF had concluded that just *homogenizing* the existing catalogues was not enough, and the approach should instead be to *re-extract the entire photometry in an homogeneous way*. Thus she embarked in this huge task, as a prerequisite for the concurrent work on Paper VI (the 1000 brightest AGN, Fotopoulou et al. 2016).

The result of her work was released to me across 2015 and 2016, and, after a few checks was released on a "fast track" (i.e. without accompanying documentation !) to the

XXL consortium on 12 February 2016. In the remainder the entire dataset or any part thereof will be referred for short as SF2015.

This report describes how the SF2015 catalogues were ingested in Milan and later used for constructing Generalized Correlation Tables and views supporting multiwave catalogues, building on the experience of the predecessor project, XMM-LSS (Chiappetti et al., 2013).

The plan of this report is as follows: section 2 summarizes the motivation for the re-extraction of the photometry and gives some hints about the way it has been done; section 3 provides a listing of the input dataset supplied by SF, while section 4 describes the procedure used for ingesting it in the Milan DB and the material which is released; section 5 gives a quick comparison with the photometric catalogues formerly contained in the Milan DB and now obsoleted by SF2015; finally section 6 deals with various identification procedures and the associated catalogues accessible via the Milan DB, section 7 describes the usage of chance probabilities and likelihood ratios to rank the associations, and section 8 describes the updates to the Java visualization tool for overlaying identification candidates on thumbnail images. Last but not least, a quick starters' guide on the best way to access the views is provided in section 9.

Note that in the PDF version of this report links like this are active URLs giving access to the original web pages via the browser of your choice. Unless a different colour coding is indicated (e.g. for tables), a **blue** font indicates material supplied by SF, a **red** font indicates example material of lesser importance, and **green** indicates material added after the semi-final draft was shown to SF and Marguerite Pierre just before Mykonos.

2. Motivation for re-extraction

The text in this section has been supplied by SF.

The re-extraction of the photometry was deemed necessary for three reasons; to salvage missing objects, provide better photometric calibration, and identify the counterparts of AGN. During the XXL-GAMA matching group

telescope or survey	camera or instrtment	filter	version	mean λ (μm)	λ_{eff} (μm)	FWHM	Vega to AB (mag)	A_V (mag)
GALEX	FUV detector	FUV	1	0.1539	0.1539	0.0228	2.363	8.310
GALEX	NUV detector	NUV	1	0.2316	0.2316	0.0796	1.770	8.736
Sloan	SDSS	u	1	0.3562	0.3562	0.0583	0.953	4.905
Sloan	SDSS	g	1	0.4719	0.4719	0.1263	-0.085	3.781
Sloan	SDSS	r	1	0.6185	0.6185	0.1150	0.155	2.723
Sloan	SDSS	i	1	0.7500	0.7500	0.1239	0.370	2.087
Sloan	SDSS	z	1	0.8961	0.8961	0.0994	0.537	1.513
CFHT	MegaCam	u	1	0.3811	0.3811	0.0654	0.402	4.690
CFHT	MegaCam	g	1	0.4862	0.4862	0.1434	-0.077	3.649
CFHT	MegaCam	r	1	0.6258	0.6258	0.1219	0.167	2.687
CFHT	MegaCam	i	1	0.7690	0.7690	0.1367	0.401	2.000
CFHT	MegaCam	i ¹	2	0.7553	0.7553	0.1571	0.376	2.064
CFHT	MegaCam	z	1	0.8871	0.8871	0.0935	0.535	1.537
CFHT	WIRcam	K	1	2.1520	2.1520	0.3256	1.837	0.366
UKIRT	WFCAM	J	1	1.2510	1.2510	0.1590	0.929	0.876
UKIRT	WFCAM	H	1	1.6380	1.6380	0.2920	1.369	0.569
UKIRT	WFCAM	K	1	2.2085	2.2085	0.3535	1.884	0.351
VISTA	VIRCAM	z	1	0.8590	0.8590	0.0979	0.525	1.586
VISTA	VIRCAM	Y	1	1.0224	1.0224	0.0926	0.613	1.212
VISTA	VIRCAM	J	1	1.2548	1.2548	0.1725	0.931	0.872
VISTA	VIRCAM	H	1	1.6466	1.6466	0.2916	1.377	0.564
VISTA	VIRCAM	K	1	2.1471	2.1471	0.3092	1.827	0.368
Spitzer	IRAC	ch1 ²	1	3.5634	3.5504	0.7444	2.798	0.163
Spitzer	IRAC	ch2 ²	1	4.5110	4.4930	1.0119	3.265	0.112
WISE	survey camera	W1	1	3.3792	3.3573	0.7935	2.670	0.179
WISE	survey camera	W2	1	4.6293	4.6065	1.1059	3.310	0.107
WISE	survey camera	W3	1	12.3338	11.8178	8.6745	4.732	0.024
WISE	survey camera	W4	1	22.2532	22.1355	4.5568	6.325	0.009
CTIO	Mosaic-2	g	1	0.4820	0.4820	0.1396	-0.077	3.686
CTIO	Mosaic-2	r	1	0.6266	0.6266	0.1399	0.167	2.681
CTIO	Mosaic-2	i	1	0.7759	0.7759	0.1414	0.418	1.967
CTIO	Mosaic-2	z	1	0.9800	0.9800	0.1286	0.534	1.499
CTIO	DECam	g	1	0.4763	0.4763	0.1493	-0.085	3.744
CTIO	DECam	r	1	0.6433	0.6433	0.1485	0.193	2.601
CTIO	DECam	i	1	0.7840	0.7840	0.1468	0.427	1.932
CTIO	DECam	z	1	0.9206	0.9206	0.1492	0.533	1.440

Table 1. Characteristics of the filters used by the various surveys (Table supplied by SF)

¹ CFHT version 2 i filter commonly referred as y

² the IRAC filters are referred as 3.6 μm and 4.5 μm

activities A. Bongiorno and M. Brusa identified cases of missing objects from the CFHT catalogs. This does not come as a surprise, as it is a well known issue of general catalog extraction (already reported e.g. for the COSMOS field in Brusa et al. 2007).

Photometric redshift estimation relies on the relative calibration of the photometric bands used. Often catalogs and images downloaded from astronomical archives suffer for calibration offsets, which can be attributed to weather conditions, and/or incorrect processing of the whole or parts of the survey. In order to identify and correct for said offsets, one must compare the colors of candidate stars against models, since the stellar colors define a narrow locus in color space. To do so, the photometric catalog of a full image is needed along with a source extension

measurement (such as the half light radius) to identify bright point like sources.

Lastly, in order to identify the counterparts of XMM sources we must rely on statistical methods to assign a probability to a source to be the true counterpart. This assignment is based on the magnitude distribution of sources in the survey. In order to have a correct probability estimation, the parent photometric catalog and the AGN candidate catalog must be produced through the same procedure, hence introducing missing objects by hand would not be the optimal way.

The properties of the filters used by the various surveys is summarized in Table 1.

3. The input dataset

The immediate result of the SF2015 re-extraction are the so-called *primary catalogues* i.e. a family of 47 bulky FITS files named `North_survey_filter_primary.fits` (28 files) or `South_survey_filter_primary.fits` (19 files), i.e. one monochromatic catalogue per survey per band.

The bulk files (compressed with `fpack`) are made available for retrieval as *data products* as described in section 4.4. The primary catalogues are used to construct queryable database *panchromatic* tables ($10''$ subsets) as described in section 4.1.

A synopsis of the available surveys, filters, files and tables is presented in Table 2.

In addition to the primary catalogues, I have received two *association files* which correspond to the multiwavelength association done by SF (and implicitly the multi-band association within each survey). These are not directly released but were used to build the panchromatic tables (see 4.1) and the GCTs underlying the views (see 6).

The primary files share an alike structure. Most of them are FITS binary tables with 43 columns. The CFHT files and the UKIDSS J and K files have 47 columns. The GALEX files have 48. The GALEX files have an additional numeric identifier in column 1, for the rest their columns 2-48 are the same as 1-47 of the other files. All other files have the same layout for columns 1-38 and 41. Columns 39-40 and 42-43 have different names but similar functions. Columns 44-47 are additional and are related with the procedures of tile merging and sub-survey merging.

The primary files (in uncompressed form) can be quite bulky (the smallest is SDSS g band with 89127 objects, the biggest is CFHT g band with 6278723 objects; and having more than 1 or a few million entries is not uncommon).

The primary files contain some redundant or plethoric information. For instance a column contains the filter, which is of course identical for all objects (and in a few cases it is also wrong, the correct filter being the one in the *filename*). Another redundant column assumes identically the values 'North' or 'South'. Two other columns define the survey and sub-survey but this is not always used consistently. Usually the survey is redundant (identical for all objects), but for instance the VISTA files store the sub-survey (VHS VIDEO VIKING) in the survey column. Sometimes also the sub-survey is redundant (identical). Finally some numeric pointers or values are in the FITS K format (64-bit integer) or in double precision (D 64-bit float) when a 32-bit quantity would suffice, making the files larger than they should.

However the primary files are provided (see 4.4) *unedited* as they were supplied, but for the `fpack` compression. The *names* of the files instead were normalized so that the *survey* name is in *all capitals* (e.g. WIRCAM not WIRcam) and the *filter* name is usually all lower case

(e.g. nuv and not NUV, w1 and not W1) unless it is a customary NIR band (Y, J, H, K). The replacement *i* filter for CFHTLS is called *y* according to the CFHTLS convention. For consistency with an earlier internal convention, the Spitzer IRAC bands are called 36 and 45 (for 3.6 and 4.5 μm), and not ch1, ch2.

The *association files* are respectively a 6697788×31 or 4143094×22 binary table. There are 3 columns more than the number of survey/filter primary files in the area. One column is a sequential (master) identifier, two columns are the α, δ coordinates preferred by SF's merging, and the remaining 28 or 19 columns are the pointers to the unique identifiers in the primary files (usually column 39 of the latter, column 44 for the 47-column files, column 45 for the GALEX files). The pointers can also be null if a given band is not used in the association.

The association files were produced for the purpose of computing photometric redshifts, therefore a requirement was that a source shall be detected in at least two bands (whatever the survey). Single band detections are present in the primary catalogues, but absent in the association files.

Since for ease of comparison with the published catalogues and those formerly in the Milan DB I needed *panchromatic files* I used as starting point (but not exclusively) the association files (preserving therefore the original SF associations) as described in the next section.

3.1. Additional material

I have ingested and I am partially releasing contextually with this report also the following tables provided by SF via the XXL AGN evolution wiki. They are named and arranged in uniformity with the SF photometric tables.

3.1.1. Photometric redshift tables

The two tables `SFNzphot` and `SFSzphot` contain the "best" photometric redshift and a handful of additional columns for the objects included in *any* of the SF photometric tables listed in Table 2, i.e. 169804 northern and 122765 southern objects. De facto, due to the procedure described below in 4, this means the objects within $13''$ from X-ray positions. The objects are linked to the SF photometric tables via the master identifier (called `id` within the Milan DB and `NUMBER` in SF's FITS files). The entire tables with all sources (bulky) and with the Probability Density Functions (very bulky) remain so far available via the quoted wiki page. In the future they could possibly be linked as data products.

3.1.2. Identification tables

The two tables `SFNident` and `SFSident` should contain the preferred counterparts chosen by SF. However, since the tables on the wiki are said to be not final, are not doc-

North							
Survey	sub-surveys	mnemonic	filters	DB table	# objects		supersedes
SDSS	–	SDSS	u g r i z	SFNsdss	13615	2936	sdssdr10
CFHT	D1 W1 WA WB WC	CFHT	u g r i y z	SFNcfht	147471	22650	w1t7 zt7 cfhtlens
VISTA	VHS VIDEO VIKING	VISTA	z Y J H K	SFNvista	63249	18610	–
UKIDSS	UDS DXS	UKIDSS	J H K	SFNukidss	41722	6164	ukidssdr10
WIRcam	–	WIRCAM	K	SFNwircam	43667	3125	wircam
IRAC ¹	–	IRAC	36 45	SFNirac	59278	1609	irac2v0
GALEX ¹	DIS MIS AIS GI	GALEX	nuv fuv	SFNgalax	24527	6157	galaxgr6
WISE	–	WISE	w1 w2 w3 w4	SFNwise	20646	16920	wise
South							
Survey	sub-surveys	mnemonic	filters	DB table	# objects		supersedes
BCS	–	BCS	g r i z	SFSbcs	74717	20082	bcsru bcslmu
DECam	–	DECAM	g r i z	SFSdecam	116446	74627	decam
VISTA	VHS	VISTA	J H K	SFSvista	32055	4432	–
SSDF ¹	–	IRAC	36 45	SFSirac	47728	5191	ssdf2v9
GALEX ¹	MIS AIS GI	GALEX	nuv fuv	SFSgalax	9948	2172	galaxgr6
WISE	–	WISE	w1 w2 w3 w4	SFSwise	15735	15124	wise

Table 2. Summary of the material in the SF2015 dataset.

The surveys are shown in the order used by SF for multiwavelength merging (but see note 1). The Milan DB presents them in wavelength order (UV, optical, NIR, IR). Sub-surveys are listed only when of some relevance. The names of the primary catalogues are built using the survey and filter mnemonics in the third and fourth column, while the fifth column gives the name of the Milan DB panchromatic table. Such table supersedes all previous photometric tables, the most recent of which are listed in the last column. The two columns under “# objects” list respectively the number of objects coming from SF’s master association files, and the “appended rejects” (single band detections, see text).

¹ SF apparently considered IRAC ch1, GALEX NUV, IRAC ch2 and GALEX FUV interleaved in this order for multiwavelength merging.

umented in detail, and are stored in a provisional dropbox location, I am currently **not** releasing such tables, but just using them for the comparison with my own tables in section 7.5.3). They are however visible as they stand using the tool described in section 8.

The identification tables contain a subset of the columns in SF’s tables, while they contain all objects, in terms of *exactly one entry* per X-ray source. The `seq` column provides link into the X-ray table `Xseq`, while the `id` column provides link into the SF photometric tables. I omitted all the columns duplicated from the X-ray catalogues `XXLN` and `XXLS`, and kept only a couple of coordinates (the nature and difference of them is undocumented), some columns related to the quality of the identification (Likelihood Ratio, reliability, flag) and some info about the spectroscopic and photometric redshift.

4. Database ingestion

My wish (for analogy with previously available photometric catalogues) has been to generate *panchromatic tables*, i.e. one database table per survey with the measurement in all bands/filters and a single position. Also for analogy with material previously available in the Milan DB, these should be *horizontal* subsets (i.e. include only the objects within the surroundings of X-ray sources), as well as *vertical* subsets (i.e. include only interesting columns).

The first part of this section describes the procedure I used to ingest the input dataset (described above in 3), while the output tables and other material are described in subsections 4.1 and 4.4.

A log of the ingestion is available on the web separately for the north and south areas (also clicking here).

- The first step for the ingestion (scripts `createphotonorth` and `ingestnorthassoc` or `createphotosouth` and `ingestsouthassoc`) stores in a semitemporary (hidden) master table the content of the *association files* for a “safe subset” (provisionally objects within 13” from the X-ray positions). During ingestion I re-ordered IRAC and GALEX pointers (they are grouped by survey and not interleaved), and handled null pointers in a convenient way (scripts `fixnullphotonorth` and `fixnullphotosouth`).
- The rationale to use a somewhat larger radius (13 instead of 10”) was to “play safe” in the case position uncertainties during merging would otherwise leave out valid combinations.
- This step reduces the entries to be considered from 6-4 millions to just 169804 (north) or 122772 (south).
- The next step (script `createpanchro`), to be run on each survey, i.e. for each database table listed in Table 2, and creates the panchromatic tables with a common layout (see 4.1 below) and initializes their content from

the semitemporary master.

This means copying the pointers for all bands of the survey (null or not null) and the master identifier (as well as provisionally SF's α , δ) from the master for all entries where one of the survey/band pointers is not null ("pre-application of the horizontal cut").

- The next step (script `ingestpanchroband`), to be run on each survey and band, requires that the primary catalogues have been renamed (or soft-linked) according to the standard naming convention.

It first ingests the chosen columns from each primary catalogue in a temporary table (applying the "vertical cut"), and then fills the specific band columns in the panchro table for the objects with a master pointer already filled by the previous step ("horizontal cut").

Additionally the script stores just the coordinates (truncated to 0.01 deg) and the magnitude of *all objects* to a "band density table" to be used later for probability or likelihood computation (see 7 below and in particular 7.2.1).

In addition the script does also a correlation within $10''$ with the X-ray position, and notes the objects (presumably single band detections) which were not inserted in the panchro table because not pointed at by the master table. Their relevant data (the chosen columns) are saved ad interim to a "band reject table".

- The following step (script `insertreject`) is run for each survey and band, and internally does a loop on all *other* bands of the same survey. In a few cases it has to be run recursively.

It is not run for WIRcam which has a single band. It looks for the objects in the "band reject tables" whose position is within the canonical distance from the position in the current band for a given master identifier, and inserts their data (the chosen columns) in the empty space in the record pointed by the master identifier.

The *canonical distance* is $0.7''$ for optical and NIR surveys, and $2''$ for the other (GALEX, IRAC, WISE).

The affected sources are flagged as `supplemented`, with the indication of the supplementary band. Relatively few (from a few tens to max a few hundred) objects per survey/band are affected.

An example of this is presented below at the end of the introductory part of section 4.1.

- The following step (scripts `appendreject` and `appenduncond`) is run for each survey and group of bands. E.g. for BCS one calls `appendreject` first for bands griz, then riz, then iz and finally `appenduncond` for band z. For WIRcam only `appenduncond` is called for band K.

These scripts deals with the objects in the "band reject tables" not disposed of in the previous step. Such objects can be truly single band detections (not considered by SF in her association files), to be *unconditionally appendend* to the panchro table. Or can be

close within the canonical distance to the reject in another band.

As a further simple example consider IRAC with two bands. `appendreject` is run on the first band ($3.6\mu\text{m}$). It first appends the updates of $3.6\mu\text{m}$ objects close to a $4.5\mu\text{m}$ object, then inserts the data for such $4.5\mu\text{m}$ objects in the relevant $3.6\mu\text{m}$ entry, and finally unconditionally appends the remaining $3.6\mu\text{m}$ -only objects. Then for the *last* band (here $4.5\mu\text{m}$) one calls `appenduncond`.

The affected sources are flagged as `appended`, with no indication of the band, as all bands, if more than one, are "new". The master identifier does not point to anywhere in the master association file (these objects are additional to SF's associations), but assumes conventionally a *negative value*.

- For convenience some of the above scripts works by repeated application of the script to chunks of max 0.5 million entries (but this is transparent to the result).
- The final steps are not logged nor handled by a script. They can be considered provisional. They are applied to each panchro table.
- First of all the "official coordinates" (columns named `ra,dec1`) for the *appended objects* are set to the coordinates of the first non-null band in customary wavelength order.
- Then a "spurious" flag (with an additional band indication) is set for *all* cases where a (non-null) band coordinate differs from the official coordinate by more than the canonical distance ($0.7''$ or $2''$).

Note that at this stage the official coordinate of the objects in a panchro table for a given survey, coming from SF's master associations *might* be a coordinate taken from *another* survey ! Usually the spurious flag will be set only in conjunction with the supplemented flag, and pointing to the same band, giving rise to combinations like `supplemented,+i,spurious,-i`

- A further step *resets* the official coordinate for the objects from SF's master associations (positive *id*) to the coordinates of the first non-null band *in the specific panchro table* in customary wavelength order. If the coordinate changes a "revcoord" flag is set. Of course "revcoord" is set for very few objects for the first surveys in SF's processing order (e.g. just 27 cases for SDSS, or 1 for BCS), and is set for most objects in the other surveys (e.g. in 14672 cases for CFHT, or 70136 for DECam, where obviously the preferred coordinate derived from the earlier survey).
- Finally the "spurious" flag is recomputed (as above), using the newly revised official coordinate.

Note that the panchromatic tables as produced by the above procedure are at this stage *not fully homogeneous*. They currently include SF associations within $13''$ from X-ray positions, but "appended rejects" only within $10''$.

4.1. The panchromatic tables

There are 14 tables (8 in the north and 6 in the south) as listed in Table 2. Their name is respectively *SFNsurvey* or *SFSsurvey* where the survey name is in lower case. These tables were released to the XXL workspace on 12 Feb 2015, and appear under the categories *North photometry (SF 2015)* and *South photometry (SF 2015)* in wavelength order from UV to IR. The categories appear just after all the X-ray categories, since these tables are intended to supersede previous (non-homogeneous) photometry tables (these are currently still offered).

All the SF2015 ables share a common layout.

There are 6 common columns, and blocks of 13-15 columns for each band. The choice of columns might change (I could remove redundant ones, not advertise less important ones, or add further columns when requested); consult the "table help" within the DB, clicking on the table name. Here I give only some general information.

The band-specific columns have names made from a common pattern and suffixed by the band name. So the pointer-identifier in band u is called *idu*, and the one for band z *idz*, or the aperture magnitudes for bands J and K are called *magJ* and *magK*.

Some columns like the α, δ in the band, or the E(B-V) in the band have "hidden" names (e.g. *_rau*, *_decu*, *_ebvu*), might not be advertised or could be retired in the future.

All bands have currently a choice of 13 columns (identifiers, positions, magnitudes, some flags or extension parameters). If the *idband* column is *null*, it means the object was not detected in the band. All other columns are set to zero. For some surveys there is a *tileband* column, and for some a *surveyband*. They report the tile id, or the sub-survey, relevant to the particular detection. This info was included only when useful for the comparison with the old data, and could be retired in the future.

The common columns include the usual sequence identifier *seq*, the record timestamp, the official (revised) coordinate *ra, decl*, the master identifier *id* (which is positive for objects from SF's master associations, and in this case it is the same *for all surveys*, and negative, equal to *-seq*, for appended rejects), and the *sflags*.

The latter is a *mysql SET*, i.e. it can assume a comma-separated list of values drawn from the following list:

- *supplemented* indicates one of more of the band detections was not listed in SF's master association but was recovered from the band rejects
- *+band* e.g. *+u* is present alongside with the previous flag to indicate the affected band
- *appended* indicates the entire record was not listed in SF's master association and was taken from the band rejects (these have also a *negative id*)
- *spurious* indicates the band coordinate differs by more than the canonical distance from the official coordinate (SF's original master or revised)

- *-band* e.g. *-u* is present alongside with the previous flag to indicate the affected band
- *revcoord* indicates the official coordinates taken from the first present band were revised w.r.t. the original ones in SF's master association

The simplest way to query a given flag is set is a condition like *sflags like '%spurious%'*. Users familiar with *mysql* surely know other ways (*find_in_set* or numeric codes).

Let us take a simple example of a reject in band g for SDSS. The object with *SFNsdss.id=17068* was present in SF's master association as an *uriz* detection. Its official coordinates were those of the SDSS u band (and are confirmed as such, *ra=_rau*, *decl=_decu*). The distances of the *r,i* and *z* detections from the u band are respectively 0.55", 0.42", 0.38". There was a g reject (*idg=160749*) whose distance from the u position is 0.75", which justifies the rejection. However the g position differs only by 0.21", 0.34", 0.37" from the *riz* positions. Therefore the g reject is *inserted* into *SFNsdss.id=17068* and assigned *sflags=supplemented,+g,spurious,-g*.

4.2. Additional material

4.2.1. Photometric redshift tables

The input is provided by the *North_photoz.fits* and *South_photoz.fits* retrieved from the AGN evolution wiki. A first step (script *createzphot*) creates empty tables *SFNzphot* and *SFSzphot*. A second step (script *ingestzphot*) ingests the selected columns for all objects (in chunks of one million sources at a time) in a temporary table, and moves the objects whose *id* is present in the hidden master association tables into *SFNzphot* or *SFSzphot*.

4.2.2. Identification tables

The input is provided so far by the preliminary FITS files stored on dropbox. A simple script *createnorthsfident* or *createsouthsfident* (the differences are due to the hardcoded names of the provisional input files and to a difference in the number of columns) creates the table and ingests the chosen subset of columns.

4.3. Correlation tables

With the release of the SF2015 photometry tables, also the following correlation tables are offered.

4.3.1. Correlation with X-rays

All SF2015 tables are correlated with the complete band merged X-ray tables (*north33*, *south33*), and with the dependent catalogues/views (*XXLN*, *XXLS*). They are also correlated with the separate N and S sections

Survey	# old	# new	1:1 common	old ambiguous	new ambiguous	old lost	new lost
North							
CFHT	128970	116021	94887	966	296	32587	24018
SDSS	20794	13081	5615	20	1883	10891	5596
UKIDSS	17313	32545	15059	90	32	2106	32545
WIRcam	51358	32540	24551	7	1	26798	9211
IRAC	34265	41500	25728	4015	517	4110	8961
GALEX	42974	23016	6040	5524	8664	16725	7776
WISE	12055	33020	6586	1129	0	? 14678	24078
South							
BCS	32052	68836	30559	448	66	916	39014
DECam	101774	148209	67071	4207	9376	17173	73433
IRAC	82882	37548	29581	3075	372	49837	3373
GALEX	10647	9586	4470	486	2089	? 44272	2636
WISE	10338	27318	5595	785	0	? 16013	20145

Table 3. Statistics of the comparison between old and new photometry.

(agn1000n, agn1000s) of the 1000 brightest AGN catalogue (Fotopoulou et al., 2016).

Ticking on *one* of the SF2015 photometry tables and *one* of the X-ray tables will popup in the right column a correlation table labelled "corrected position 10 arc-sec". This will remain valid irrespective of the presence or deletion of the 13" objects in the photometry tables, and speed up anyhow any query for a radius smaller than 10". I am currently not providing a correlation table at 6", just emulate one, using the 10" one, with the condition $\text{dist}(Xra, Xdec1, SFxxx.ra, SFxxx.dec1) * 3600 < 6$ ¹

4.3.2. Correlation among SF tables

A side effect of the fact all SF tables share a common *id* with the master association, is that the "By identifier match" correlations, which popup ticking on *two or more* SF tables of the same area, are operational.

Such correlations, although supported within **DART**>, are normally not operational within XXL, because XXL tables do not have the concept of an identifier common among more tables. They do operate within SF tables *but most likely not quite like users expect*.

For instance if you tick on two SF tables and on the "By identifier match" correlation, you will be able to see all the columns in both tables for all objects which are detected in both surveys. If you tick on three or more, you will be able to see only the objects which are detected in *all* the surveys selected. Objects detected in one survey only, or anyhow not detected by at least one, *are not shown* (technically, "By identifier match" implements an INNER JOIN, while the XXL customary views implement a LEFT JOIN.

¹ The condition shown is for the catalogue/views; for the complete band merged table use something like $\text{dist}(\text{north33.ra_corr}, \text{north33.dec_corr}, \text{SFNxxx.ra}, \text{SFNxxx.dec1}) * 3600 < 6$

4.3.3. Other correlations

Views allowing to see all photometry tables together inclusive of objects not detected by all surveys will be provided only for the candidate counterparts (inclusive of the rejected ones) via the "multiwave views" described in 6.3 below. The panchromatic tables will be member table of the underlying GCTs.

4.4. The data products

Any valid query on an SF table will also give access to the *bulk primary catalogues* (containing all columns and all objects) as data products. From the "QUERY RESULTS" page, click on "Retrieve all objects related files" and you will be able to download the associated "primary catalogue (packed FITS)".

The files will have names in the form *North_survey_filter_primary.fits.fz*. They are binary tables compressed with the HEASARC utility *fpack*, using the "binary table tiled compression" in process of being incorporated within the FITS standard. Once retrieved they *must* be decompressed using the *funpack* utility before being used.

fpack and *funpack* (if not already available within your *ftools*) can be retrieved from <https://heasarc.gsfc.nasa.gov/fitsio/fpack/> and built according to instructions in one minute.

5. Comparison with older data

I have run some standard comparisons between the content of the SF2015 tables (hereafter *new*) and the corresponding tables from public or private sources present previously in the Milan DB (see first entry in last column of Table 2, hereafter *old*). *I compare a (representative) subset and not the complete (bulk) catalogues*, i.e. what ingested in the DB. The comparison has a minor complication due to the fact the previous tables in general include

objects within $10''$ from X-ray positions, while SF2015 tables were provisionally built as a mixture of $13''$ objects from the master association and $10''$ rejects. One should restrict to $10''$,

What I expected was to find most sources detected in both analyses, at close positions, with similar magnitudes, eventually with a nearly fixed offset due to the different magnitude definition, the homogenization to AB magnitudes, the correction for galactic extinction, etc. I expected possible deviations at faint end (different "sensitivity") and at bright end (I actually hoped a better handling of saturation would overcome the problem of "bright uncatalogued sources", clearly visible on images but absent in the old catalogues). The situation is generally more complex, and not unlike the differences found between concurrent analyses of the previous data (e.g. for BCS), or between different versions of the same survey "standard" analysis.

The typical comparison procedure between old and new involves the following steps:

- do some basic counts
- generate a direct (old vs new) and reverse (new vs old) correlation table within the canonical radius ($0.7''$ or $2''$). Being this a LEFT JOIN the common objects will be the same, while the "lost" objects will be present only in one of the two correlations.
- examine *the 1:1 common objects*, those for which one old object matches exactly one new object
- for these check whether they are detected in the same tile (where applicable)
- for these check the matching distance
- for these check whether they are detected in the same bands
- for those actually detected in the same band, compare the magnitudes, also in graphical form, and verify whether deviations concentrate in some magnitude interval
- examine the *ambiguities*, i.e. when one old object matches more than one new, or one new more than one old
- examine the *old lost* or *new lost* objects, see if they have some special concentration with respect to the bands they are detected in, or our own flags, or whether they concentrate in some magnitude interval.

In doing the check one has to take into account that the way different survey processed the old data may be different among them, and different from the SF procedure. This may affect things like tile overlap handling, or generation of panchromatic information. In particular for the SF tables I used the $3''$ aperture magnitudes (columns named *magband* in the database, not columns named *autoband*), while for the older data I used whatever default magnitude was chosen at the time, so I might be comparing apples and oranges (e.g. *mag_auto*, or Vega magnitudes instead of AB ones, or whatever).

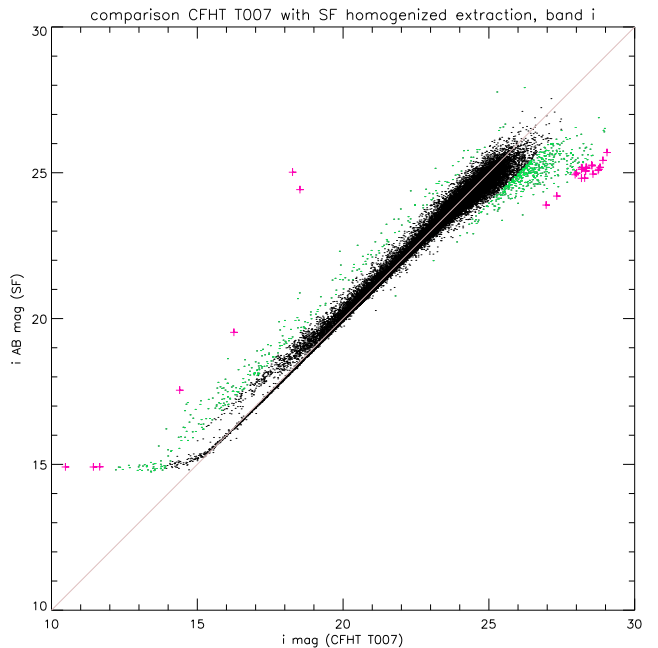


Fig. 1. Comparison between old and new magnitudes for common objects in the *i* band of the CFHT data

A summary of the essential statistics of the comparisons is shown in Table 3.

The present report shows some samples of the old-new magnitude plots. A full collection, including also histograms of the magnitude differences, is available following this link In such plots the black dots *usually* indicate magnitude differences $|old - new| < 1$, green dots $|old - new|$ between 1 and 3, and magenta crosses $|old - new| > 3$ unless otherwise stated. The pink diagonal indicates the locus of equal old and new magnitudes.

5.1. CFHT

The *SFNcfht* table, as old *w1t7*, contains both CFHTLS W1 as well as our own ABC fields. In addition *SFNcfht* contains also the deeper D1 field, not included in the old table. Due to its depth, CFHT can be assumed as the reference optical data in the North, and for this reason was examined first.

The old and new analysis are of similar depth. 82% of the new sources are in the 1:1 common subset. Of these the positional match is very good (although nominally computed within $0.7''$, 96% of the matches are within $0.3''$). The choice of tile is often different (in the SF analysis an object's magnitude can derive from different tiles unlike the CFHTLS case), but this is hardly relevant.

Also the choice of bands is done differently. CFHTLS panchromatic catalogues give a magnitude (usually not undefined) in nearly all bands. Most entries are classified as "complete" (ugriz or ugriz detections). SF analysis on independent bands may not detect at all an object in some

filters. Only 32946 cases have the same band "classification" (i.e. are detected exactly in the same bands).

This is also considered a "feature" of the two analyses, and considered hardly relevant.

Fig. 1 shows an example of the comparison of the magnitude for common objects detected in such band (i in the example shown). The other bands look alike. Usually the magnitudes match with a tiny offset, and a large scatter around 25. However there are some cases which seem to align on a non-45° slope. A moderate number of cases (in green) differ by more than 1 mag, and a few tens (in magenta) by more than 3 mag. These are usually concentrated at both ends, old objects fainter than 25 which SF recovers brighter, or very bright objects, presumably saturated in the old catalogue and recovered differently by SF: a visual inspection shows they are usually bright stars, BD or HD in SIMBAD, in a couple of cases a bright galaxy (one is an MCG in SIMBAD), and in one a crowded area. **This is just the effect of the different usage of SExtractor parameters ending up with different magnitude estimation at the very bright and very faint end, which is normal. We remind that everything brighter than magnitude 17 for CFHTLS is not trustworthy.**

The ambiguous cases are relatively few and only a limited sample was examined. It is not obvious to tell which is the right association, when distances and magnitudes are also similar. A certain degree of intrinsic ambiguity is possibly unavoidable, and no particular attention was paid to similar cases for the other surveys.

The new lost objects have nothing special for what concerns the `sflags` or the band classification. The impression is that there is some concentration at weak (> 25) magnitudes, and a limited number of bright (< 20) new sources not present in old, more or less the same for all bands.

For the old lost objects a lot (95% in the u band decreasing to 38% in the z band) are in the very faint domain (> 25), and just a few in the bright.

The old and new lost do not depend on astrometric accuracy. I repeated the counts using 1.2" instead of 0.7", and the changes are very limited.

5.2. SDSS

Apparently the SF re-extraction of SDSS data is *less deep* than `sdssdr10`. Only 43% of the new sources are in the 1:1 common subset, but for those the positional match is pretty acceptable (97% within 0.5").

`sdssdr10` has all entries nominally complete (all ugriz bands), while `SFNsdss` has just 954 entries detected in all bands.

No plots are shown. They are rather noisy with a number of objects with magnitudes matching with a tiny offset, but also several deviating by more than 1 or 3 mag, and/or located on a different diagonal. Most of the larger deviations are at the bright end.

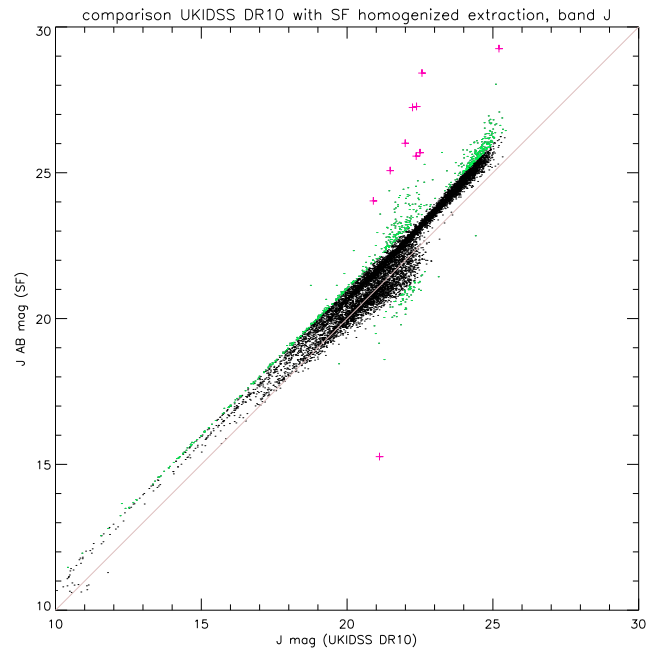


Fig. 2. Comparison between old and new magnitudes for common objects in the J band of UKIDSS data

Similarly the new lost objects show some concentration at the bright end, and no case at the faint end (> 25). While for the old lost objects a fraction is in the faint domain (more in u and less towards z), fewer in the bright domain (with the opposite trend, more in z) and most are at intermediate magnitudes.

5.3. UKIDSS

Apparently the SF re-extraction of UKIDSS data is *significantly deeper* than UKIDSS DR10. In fact UKIDSS continued taking data after DR10 which have not been merged in the deep mosaics yet. SF has used the individual tiles possibly covering a larger area. The UKIDSS collaboration is going to process all of their data for a final release, which means we could have even deeper data in some future (Bremer, private communication). As a consequence just 46% of the new sources are in the 1:1 common subset (and the ambiguities are exceptionally limited). The positional match is pretty acceptable (97% within 0.5").

Detections occur not in all bands in both old and new samples, the common cases having exactly the same classification are 9644. Fig. 2 shows the comparison of magnitudes in the J band. There is a systematic offset (becoming greater in the order JHK), but the magnitudes cluster along a 45° diagonal. A limited number of large deviations occur in the intermediate-faint domain, in the form of old magnitudes getting fainter in the SF analysis.

The new lost objects show no concentration at extrema but at intermediate magnitudes, and the old lost objects are sort of uniformly distributed.

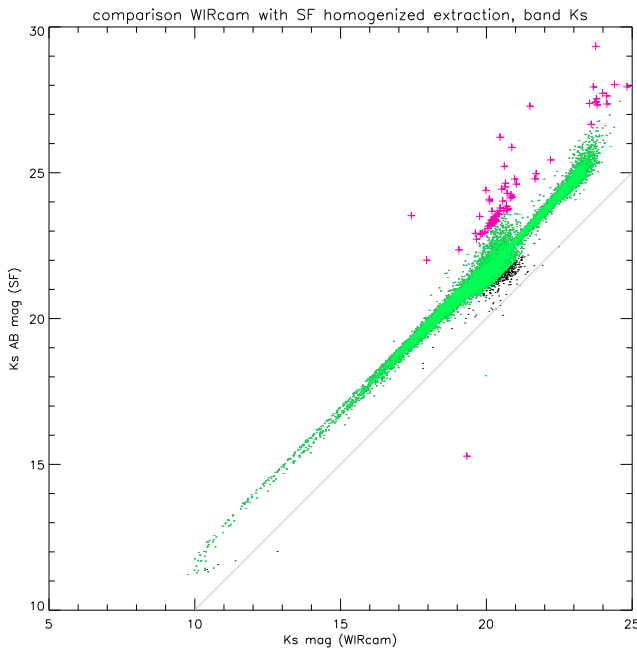


Fig. 3. Comparison between old and new magnitudes for common objects in the WIRcam data

5.4. WIRcam

Apparently the SF re-extraction of WIRcam data is *less deep* than the old one. Anyhow 75% of the new sources are in the 1:1 common subset (with almost no ambiguities, and an excellent positional match (99% within $0.2''$).

Similarly 99% of the magnitudes match within 1 mag, although Fig. 3 shows a pretty well appreciable systematic offset, due to the AB/Vega correction (see Table 1). As for UKIDSS (which is also in the NIR) a limited number of large deviations occur in the intermediate-faint domain, in the form of old magnitudes getting fainter in the SF analysis.

30% of the new lost objects are bright (< 20), the rest are normal, while old lost objects prevail at intermediate magnitudes.

5.5. BCS

SFSbcs can be used as reference optical data in the South. The comparison with the old data is made on `bcsru` as Rutgers-processed images were used also for SF analysis.

The SF re-extraction of BCS data is definitely *significantly deeper* than the Menanteau Rutgers analysis. As a consequence just 44% of the new sources are in the 1:1 common subset (plus some ambiguities). The positional match is acceptable (96% within $0.5''$).

Detections occur not in the same tile in old and new samples, also because SF allows to mix different tiles in different bands. However for 29623 the same tile has been

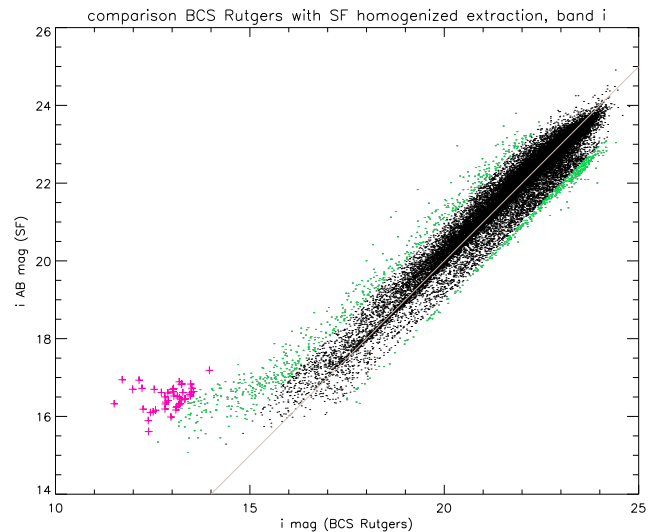


Fig. 4. Comparison between old and new magnitudes for common objects in the i band of the BCS data

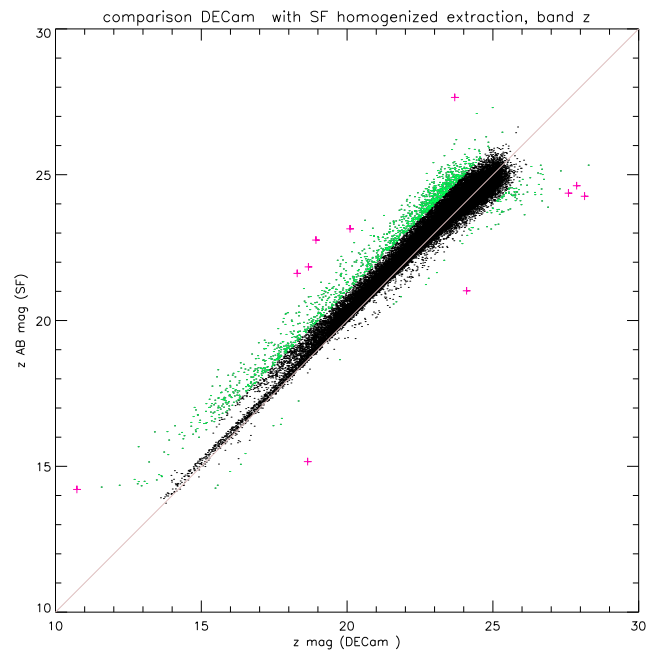


Fig. 5. Comparison between old and new magnitudes for common objects in the z band of the DECam data

used at least in one band, only 936 use exclusively an adjacent tile.

Similarly for the magnitudes. 27813 old sources are "complete" (detected in `griz`), while only 16545 SF sources are complete. 17420 have anyhow the same classification. Fig. 4 shows the comparison of magnitudes in the i band.

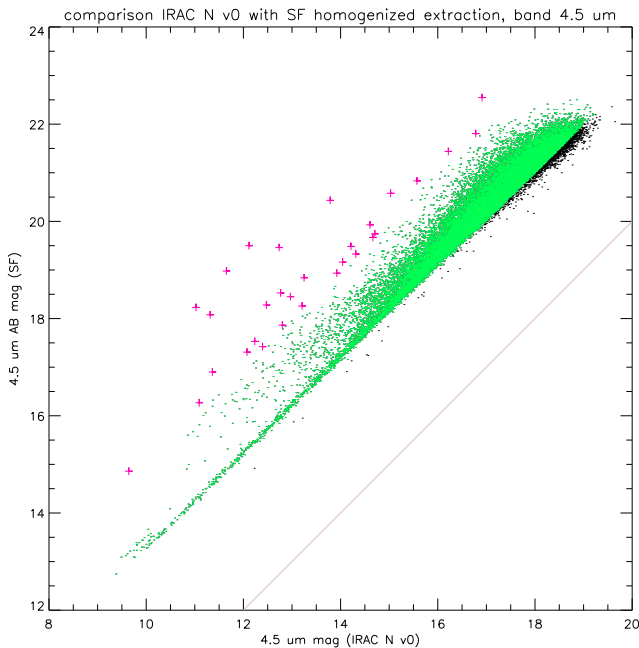


Fig. 6. Comparison between old and new magnitudes for common objects in the $4.5 \mu\text{m}$ band of the IRAC *north-ern* data. Note that green and magenta colours correspond here to magnitude differences greater than 3 and 5 magnitudes.

It is not unlike what occurs for CFHT, with old bright objects detected as fainter by SF, and old faint (sometimes absurdly faint) sources detected in the intermediate range by SF (this occurs in grz, but there are no faint objects in i). The new lost objects concentrate at intermediate magnitudes, fewer weaks and more bright towards long wavelengths. Also old lost objects are mainly in the middle.

5.6. DECam

The SF re-extraction of DECam data is *somewhat deeper* than the old analysis. Just 45% of the new sources are in the 1:1 common subset (plus several ambiguities). The positional match is acceptable (95% within $0.5''$).

Concerning tile detections, for 66820 cases the same tile has been used at least in one band, only 251 use exclusively an adjacent tile.

Similarly for the magnitudes. "Complete" detections (detected in griz) do not prevail in either old (26761) or new analysis (15466), but a good deal (44285) have anyhow the same classification. Fig. 5 shows the comparison of magnitudes in the z band. It is not unlike what occurs for CFHT, with rather few large deviations and a hint of a secondary diagonal. There are more old faint objects brighter in new in the i band and almost none in z.

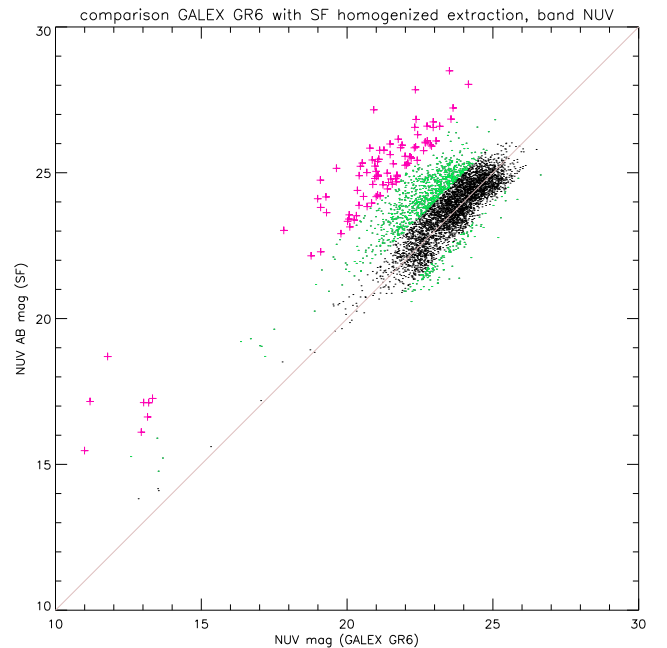


Fig. 7. Comparison between old and new magnitudes for common objects in the NUV band of the GALEX *northern* data

The new lost objects definitely concentrate at faint magnitudes, but for band z. For the old lost objects, about 50% are in the faint domain.

5.7. IRAC

Old IRAC data in the North were private data (`irac2v0`, while in the South they came from SSDF. Comparison with new data occurs separately for North and South, however with rather similar results.

The main difference between the north and south cases is that in the north the SF analysis is *deeper* than the old one, while in the south the old (SSDF private analysis extending to a 2σ level) was deeper! The positional match is decent (86radius of $2''$).

Fig. 6 shows the case for $4.5 \mu\text{m}$ in the north. All cases are similar, they show a match with a large systematic offset (and for this reason the colour coding has been changed), smaller at $3.6 \mu\text{m}$ than at $4.5 \mu\text{m}$, and with few occasional large deviations.

New lost objects are at intermediate magnitudes. Old lost objects in the south are both at intermediate mags and at the bright end (but the definition of bright is different in the IR from the optical). The north is similar but less exasperated.

5.8. GALEX

For GALEX and WISE the old data were in a single table for North and South data. This is taken into account com-

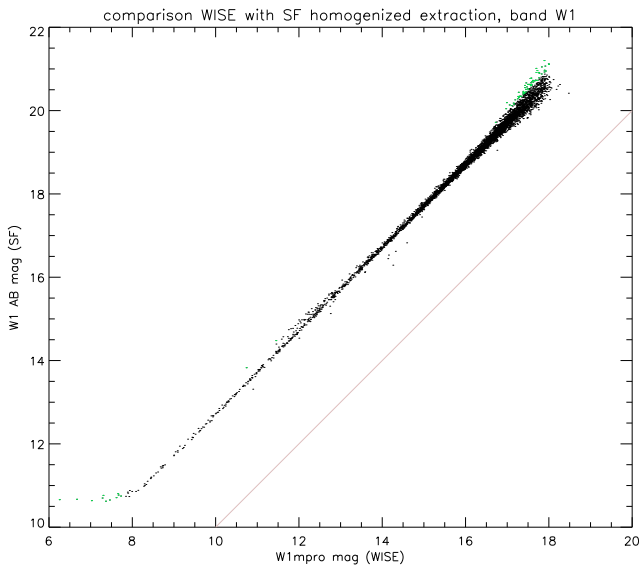


Fig. 8. Comparison between old and new magnitudes for common objects in the W1 band of the WISE *northern* data

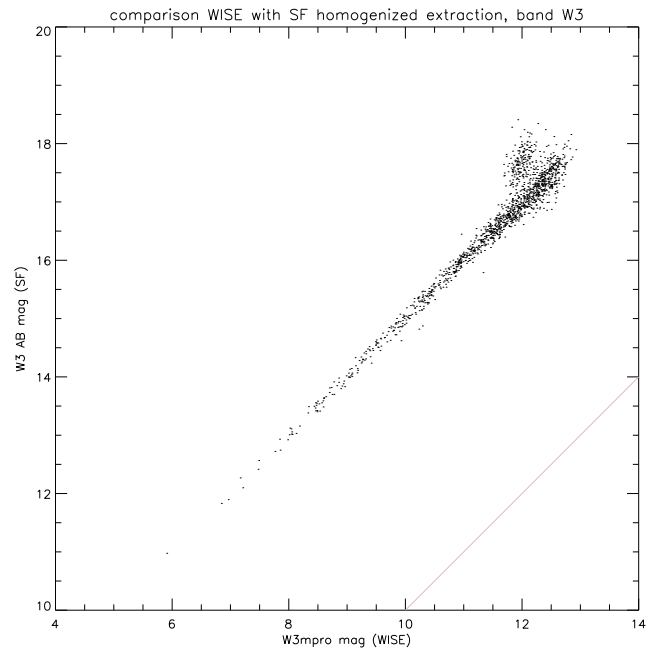


Fig. 9. Comparison between old and new magnitudes for common objects in the W3 band of the WISE *southern* data

paring separately the new data in the two areas, which are however rather similar.

The analysis by SF is *somewhat less deep* than GALEX GR6 (less obvious in the south where both are shallower than in the north anyhow).

Fig. 7 shows the magnitude difference in the NUV band for the north (with the usual colour scale). There is usually an acceptable match with a large scatter but also with a significant number of outliers. Note also that the common cases are relatively few.

Lost objects (old and new) are concentrated at intermediate magnitudes.

5.9. WISE

The analysis by SF is *definitely deeper* than WISE AllWISE both in the north and in the south. Just 20% of the new detections are in the 1:1 common subset !

Fig. 8 and Fig. 9 show respectively the magnitude difference in the W1 and W3 bands for the northern area. Considered the large systematic shift (increasing towards longer wavelengths), due to the AB/Vega correction (see Table 1), *no colour coding* is applied. Usually the points align with some scatter around a 45° diagonal (bands W1 and W2) with an additional "secondary blob" towards faint magnitudes (bands W3 and W4).

Note that the definition of magnitudes is rather different (for old data the so-called "mpro" instrumental profile-fit Vega magnitudes were used). Anyhow the lost objects concentrate at bright magnitudes, whatever defined.

6. Towards identification

The goal is to have one *multiwave catalogue* (or more) with the counterparts or candidate counterparts in the various bands related to the XXL or XXLS X-ray sources.

These catalogues shall be released in form of *VIEWs* (see 6.3). The preliminary step to create such views is the generation of the underlying *GCT* (generalized correlation table), which is described in 6.1. The next step is the assignment of some form of *ranking* to the candidate counterparts (see 7), and finally one could make the views (see 6.3). The process can be assisted by an update of the existing visualization tool (see 8) to overlay the counterpart positions over optical (or other) thumbnail images.

6.1. Creating the GCTs

For both the north and south areas, I created two alternate GCTs: one based straight on SF's associations, and one based on based on my incremental procedure already used for the XMM-LSS and for previous exercises for XXL, as described in Chiappetti (2014a) and Chiappetti (2013).

I remind that a GCT is a (hidden) database table with the following columns:

- a *seq* (record number), identifying the *counterpart set*
- a number of columns with the name of the *member tables*, they are pointers into the member tables themselves (i.e. the *seq* in the member)
- some *standard auxiliary columns* inherited from the X-ray-only GCT and used to build the X-ray source

catalogue name. Optionally there could be pointers to previous versions of the X-ray catalogues (XLSS pointers).

- a number of *ad-hoc auxiliary columns* among which the *ranks* of the counterpart sets.

All GCTs have as first 3(4) members the band-merged X-ray table (`north33` or `south33`), the individual band X-ray tables (so called b and cd tables), and eventually a duplicate of the merged table to keep track of nearby sources.

All GCTs considered here have as further members *only* the SF photometric tables (so 8 SFN in the north and 6 SFS in the south).

6.1.1. Using SF associations

As a preliminary step, I have converted the semitemporary master association tables described above in 4 into two provisional GCTs `glorsfn` and `glorsfos`. These have exactly the same number of entries (N: 169804; S: 122772) as the corresponding association table. They have only the SF tables as members (there are *no* X-ray members). They inherit as auxiliary columns the id and α, δ coordinates from the masters. In addition they have a flag telling whether the object is within 10'' or only 13'' from an X-ray source.

The difference between the masters and these GCTs is that the masters contain the id's in all bands of each survey, while the GCTs contains the seq pointer to the SF photometric tables (*the latter* contains the individual band id's).

So far `glorsfn` and `glorsfos` are not used for any purpose other than input to the next steps.

The actual GTCs with the X-ray to SF association are called instead `glorsfn` and `glorsfs` and are initialized from XXLN or XXLS, as empty tables with 14168 (N) or 11888 (S) entries, with the X-ray members filled, and the SF members empty.

Then I copy the member pointers (all of them) from `glorsfn` or `glorsfos`, for objects which are in the correlation table between X-rays and each member table (thus within 10''), taking care not to insert twice the counterparts with same SF id, and duplicating as many (different) records as needed for each X-ray source as different counterparts it has.

At the end `glorsfn` has 70854 entries for 14168 X-ray sources, of which 135 remain *completely unidentified* (nothing within 10''). `glorsfn` has 49144 entries for 11888 X-ray sources, of which 166 unidentified.

It is verified that they correspond to all objects in `glorsfn` or `glorsfos` flagged as 10'' and not associated to X-ray positions rejected by the overlap removal procedure.

6.1.2. XLSS-style associations

The alternate `glorlcn` and `glorlcs` are generated with an incremental procedure like the one used for XMM-LSS. Refer to Chiappetti (2014a) and references therein.

There are two differences with respect to the way the procedure was used so far:

- so far I had used or forced a correlation with X-ray positions within 6'', now I use the native correlation within 10''
- in the past I used a variable (user supplied) matching radius between any two non-X-ray tables. Now I use exclusively 0/7'' or 2'', consistently with SF usage (optical-NIR, and anything else)

The GTCs `glorsfn` and `glorsfs` are initialized from XXLN or XXLS, as empty tables with 14168 (N) or 11888 (S) entries, with the X-ray members filled, and the SF members empty, as customary.

Then I proceed incrementally. For the first member table I insert the pointer into the X-ray sources closer than 10'' (from the pre-built correlation table). For the next member, if the object pointed is closer than 10'' to an X-ray object, and closer than the canonical distance to any previously inserted counterpart, the pointer is inserted in the existing record. Otherwise a new record is appended (with the pointers to the previous tables being null).

For the southern area, the order followed is the natural one consistent with SF: BCS, DECam, VISTA, IRAC, GALEX and WISE.

For the northern area, I was uncertain whether to start as SDSS then CFHT, or instead CFHT then SDSS (CFHT is deeper). So I did both and compared. I refer to the former case as "base" and to the latter as "clone". I stopped after the first two steps. Both the base and clone case have 281 X-ray sources so far unidentified, and 76053 identified in the same way (of these 60393 base are identified identically to `glorsfn`, hereafter SF for short, as well as 60406 clone: 60385 are identified in the same way in all three cases).

At the end one remains with just 35 X-ray sources, with 39 base SDSS:CFHT counterparts (8 in SF) and 40 clone (21 in SF) which are different. I looked at them, and considered as bonus points the correspondence with SF and the fact to have counterparts in both surveys. On these grounds the clone arrangement looks preferred.

So the final order used for the north is CFHT, SDSS, VISTA, UKIDSS, WIRcam, IRAC, GALEX and WISE.

`glorlcn` has 124112 counterparts for 14168 X-ray sources, of which 18 totally unidentified. `glorlcs` has 118667 counterparts for 11888 X-ray sources, of which 21 totally unidentified.

Incidentally, a posteriori I inspected visually the 35 cases mentioned above with the tool described below in 8. About half the cases show definitely one pointlike object on the optical image, in some cases the affected object is

also rather far, and is unlikely to be later confirmed as counterpart. Some of those definitely-one-object are however strange, for instance for Xseq 208394 and 212453 the image shows one object but SFncfht finds two, 208516 is similar but two SFNvista, 220990 is also possibly ambiguous, 223847 is in a crowded field, and 220990 could be perturbed by the spikes of a very bright star further south.

5 cases are also one object, but this is a bright star where photometry can be confused: 202837, 212773, 214147, 217045 and 219058. Other 7 cases are a bright spiral galaxy, sometimes catalogued: 207440, 207444 (galaxy with possible pointlike enhancement on edge of optical FoV), 217662 (edge on), 218114 (aka VV 543), 218847 (aka MCG-01-07-007), 219239 and 226335.

The remaining cases are likely blends of 2-3 very near objects: 208102, 217697, 220956 and 222909. And the following peculiar cases: 201473 (a linear group of 3 sources), 203547 (3 interacting galaxies ?), 204759 (binucleated, and both nuclei observed independently by GAMA), 27198 (also binucleated).

6.1.3. Comparing GCTs

It is not surprising that the "LC" GCTs have more candidate counterparts than the "SF" GCTs (because of the "appended" single band detections). Also bear in mind that considering these candidates could be excessively conservative (they are within 10" and likely to be later rejected). Anyhow it is appropriate to compare the two couples of GCTs.

A first way is just to count the entries detected in which survey. Such statistics of the "classes" shows for instance that the most populated class in the south is "DECam only" (-D----) with 49200 occurrences in LC but 11840 in SF. Then LC has 10965 BCS only and 9658 WISE only (but these are respectively just 574 and none in SF). The next frequent case is comparable ("BCS and DECam" with 9510 LC vs 10014 SF). And so on.

In the north the most numerous class is "CFHT only" (30339 LC vs 17661 SF), then LC has 12320 WISE-only, 10093 VISTA-only (much less numerous in SF) and 7127 UKIDSS-only (in the third position for SF, while the second position is for CFHT-and-VISTA which is fifth for LC). And so on.

Now, starting from the southern objects, if one looks at the objects which have the same identification (building an "unid" or unique identifier concatenating the seq's in the various catalogues), excluding the few totally unidentified cases already mentioned, one finds that 41336 glorlcs counterpart sets are present identically in glorsfs (84% of 49144). All these (same unid) are called "direct matches".

What remains ("no-direct-matches") can be complete no-matches, or partial matches (same counterpart in some surveys, different or none in other). One starts with 7663

no-direct-match in SF and 76780 in LC. For LC almost all X-ray sources (96%) have some no-direct-matches, but only 4937 (41%) have one in SF. 6540 X-ray sources with 36138 counterparts are no-matches (present in LC but not in SF at all). There are *none* of the reverse case (SF counterparts with nothing else in LC).

So one is left with 40642 cases for the same 4937 X-ray sources. Of these 5019 have the same BCS counterpart, 5543 have the same DECcam counterpart, 3628 have the same VISTA counterpart, 6506 have the same IRAC counterpart, 1782 have the same GALEX counterpart and 3534 have the same WISE counterpart. The total of the partial matches are 12403 obviously for 4937 X-ray. 4747 X-ray sources have the remaining unassigned cases (all in LC).

Deleting the cases with no correspondence one is left with 7663 SF and 12403 LC counterpart sets for 4937 X-ray sources, which can be split according to the fact they have a single or multiple counterparts in one or both of LC and SF.

I have done an inspection by sample in the south, using the tool described in 8, on some of the discrepant cases. Since I did that first for the north, as said a few paragraphs below, I make reference to the terminology down there. The physiological edge effect of LC being limited within 10' shows up e.g. for Xseq 208386, 211360, 216276. A different physiological effect is represented by Xseq=210854 where LC adds a DECcam counterpart to a common IRAC/GALEX/WISE set (this being a true single band detection, ignored by SF "by construction"). A further physiological effect is represented by Xseq=208657. In this case LC has a BCS/VISTA/etc. association, where SF associates the same VISTA object to another BCS object. There are actually two VISTA objects very close, but different, and two BCS objects also very close (but the image is not double, can be just an elliptical). The two BCS are formally a separate gri and z detections but perhaps they should be band merged in a single one.

The few cases (just 7) detected in all the same surveys but with some different counterpart are all due to the 10' edge physiological effect (201434, 205973, 207197, 216475, 222906, 226348 and 226983).

There are the usual cases of "missed survey mergers". For instance for Xseq=200701 SF has a separate BCS/DECcam and VISTA counterpart sets, which LC correctly merges into one (here the reason is that BCS/VISTA distance is 0.76" while DECcam/VISTA is 0.37"). For Xseq=200701 (an edge-on spiral with an ATCA resolved radio counterpart), SF misses the merger of a GALEX object whose position is slightly offset. Conversely 200014 and 200056 are cases in which SF correctly merges two LC "missed mergers". Xseq=210014 is a combination of missed merger (LC BCS vs DECcam/VISTA correctly merged by SF) and a physiological case in which LC chooses a BCS/DECcam combination, and SF another BCS/DECcam (with the same BCS):

there are two very close distinct faint DECAM, and the due to the faintness the image is unclear.

$Xseq=207115$ may look a triple missed merger, but is possibly confused by two close WISE objects in a group of 3 near optical sources. In fact the two WISE objects are the **same** object, assigned by SF's association procedure to two different ids (274251 and 1947915) with exactly all the same four $idw1$, $idw2$, $idw3$, $idw4$.

Finally a case like $Xseq=200015$ is a mixture of the preceding cases. SF has 3 counterpart sets, LC has 5. One of the LC cases is a single IRAC object which SF physiologically associates to VISTA and WISE beyond $10''$. But the other two SF 1798:1680:1685:1620:765:908 and 0:0:31253:46824:0:15160 match LC 1798:1680:31253:1620:765:908 (with a different VISTA), 0:0:1685:1620:0:908 (with the same VISTA, IRAC and WISE but no optical), and 0:0:0:0:0:15160 which has the same WISE as the second SF. Now the two VISTA objects with seq 1685 and 31253 should positionally be the same object, one is nominally detected only in H and the other in JK, so they get natively different ids in SF association procedure but should be band merged into one. Instead the two WISE entries with seq 908 and 15160 are the **same** object, assigned by SF's association procedure to two different ids (132461 and 3932169) with the same $idw3$, $idw4$ in the two bands where detected.

Going back to the north, for the objects which have the same identification, excluding the few totally unidentified cases already mentioned, one finds that 60162 glori1cn counterpart sets are present identically in glorsfn (85% of 70854), i.e. "direct matches".

One starts with 10575 no-direct-match in SF and 63585 in LC. For LC almost all X-ray sources (95%) have some no-direct-matches, but only 6307 (44%) have one in SF. 7137 X-ray sources with 23665 counterparts are no-matches (present in LC but not in SF at all), and again no reverse case.

So one is left with 39920 cases for the same 6307 X-ray sources. Of these 1438 have the same SDSS counterpart, 7496 have the same CFHT counterpart, 5532 have the same VISTA counterpart, 2970 have the same UKIDSS counterpart, 3634 have the same WIRcam counterpart, 8619 have the same IRAC counterpart, 4761 have the same GALEX counterpart and 5278 have the same WISE counterpart. The total of the partial matches are 17688 obviously for 6307 X-ray. 5943 X-ray sources have the remaining unassigned cases (all in LC).

Deleting the cases with no correspondence one is left with 10575 SF and 17688 LC counterpart sets for 6307 X-ray sources, which can be split according to the fact they have a single or multiple counterparts in one or both of LC and SF.

I have done an inspection by sample, using the tool described in 8, on some of the discrepant cases. A com-

mon, physiological occurrence is represented by cases like these. Consider $Xseq=201935$, where LC has an isolated WISE-only counterpart, while SF assigns the same to a CFHT/IRAC/WISE counterpart. LC "misses" those two CFHT and IRAC, because they are at more than $10''$ from the X-ray position while WISE is at $0.5''$. Similar for a GALEX-only for $Xseq=210372$. In general the fact that SF unid's are "longer" (have more surveys with non-null detection) can be explained by the above physiological behaviour (SF works with all data while LC has edge effects at $10''$). This is likely to be hardly relevant for actual counterparts which are much closer to the X-ray position.

The same physiological effects applies to the very few (just 6 cases) where there are non-null detections in exactly the same surveys, but the choice in one of the survey differs. Consider 202736 where LC prefers an "appended" VISTA to one farther than $10''$; or 202876 and 204305 (same for CFHT); or 217840 (GALEX) and 220855 (IRAC) in which the object taken by LC is also closer to the other counterparts. For 203299 LC prefers an IRAC object which looks to match the SDSS, CFHT, VISTA and WIRcam positions, while SF associates such IRAC object with a fainter displaced CFHT and uses in its stead one which is not only beyond $10''$, but further displaced and not even visible in the IRAC images.

A variant of the physiological effect is represented by $Xseq=201008$ where LC has a single unid 0:12812:12124:0:7677:55803:0:0 while SF has two partial matches 0:12812:12124:0:7677:0:0:0 and 0:0:0:0:41673:55803:0:0. Here IRAC seq=55803, slightly displaced north w.r.t the other positions, is further "pulled out" to the north in SF by a different VISTA object beyond $10''$.

Other cases have to do with possible counterpart (survey) merging (two unid's of a form e.g like S:C:V:U:0:0:0:0 and 0:0:V:0:w:I:G:0 with a common VISTA counterpart could be merged into one S:C:V:U:w:I:G:0). A merger can be missed by LC (e.g. 227229, 200002) or SF (e.g. 226324 ...this one is actually probably something trickier ...apparently SF has a CFHT/VISTA/WIRcam :0:107685:45119:0:32076:0:0:0 and an IRAC-only 0:0:0:0:0:57573:0:0 which could be merged into one like LC does 0:107685:45119:0:32076:57573:0:0. The IRAC object is slightly displaced, but the merging makes sense. The point is that there is a single SFNirac.seq=57583 object in the catalogue, but the IRAC images show two objects in correspondance of two optical ones, and 57573 is a blend in between !).

A different kind of problem seems native to the master association procedure (not unlike a possible problem in the LC procedure, i.e. unid's for the same X-ray source where the same pointer to a given survey appears with the same value in two unid's). A simple example is for $Xseq=200064$. SF

has one unid 0:64830:0:0:0:30286:0:0 using IRAC object 30286. In LC such IRAC object is alone in 200064:0:0:0:0:0:30286:0:0 while it is replaced by another one in 200064:0:64830:0:0:0:50247:0:0. The point is that IRAC 30286 and 50247 are the **same object** ! I noted that because they have exactly the same 3.6 μ m magnitude. But if one looks at 30286 it has an *id36* of 1082237, and 50247 has the same *id36* (which confirms they are the same 3.6 μ m detection). In SF master association, 30286 corresponds to *id*=2844406 which is 3.6 μ m-only, while 50247 is *id*=5995417 with the same *id36* (plus a non-null 4.5 μ m detection).

A more contorted example occurs for *Xseq*=200005. SF has 3 unid's in partial match with 6 LC unid's. One couple is simple (SF has an extra IRAC object beyond 10''), but the remaining two SF (0:43762:26057:9615:21275:0:17250:0 and 0:98433:43663:18984:30586:39050:0:0) and 5 LC (0:43762:26057:9615:21275:0:11282:0, 0:0:0:30755:0:17530:11282:0, 0:0:0:0:0:52812:11282:0, 0:0:0:0:0:0:17250:0 and 0:0:0:0:0:0:23794:0) are "crossed" among themselves in a twofold way. The three GALEX objects with seq 11282, 17250, and 23794 are the **same object** with *idnuv*=1761063 but are reported in three different *ids* in the master association (872164, 1843229 and 6279682). But also the IRAC objects with seq 17530 and 52812 are the **same object** with both *id36*=1092035 and *id45*=969806, reported in two *texttids* in the master association (872164 and 6279682).

6.1.4. Secondary members

The secondary member columns, or related, are *id*, *SF?zphot*, *usno*, *simbad* and *ned*.

While *glorsf?* GCTs contain natively an *id* which points directly to the master associations made by SF, this information is not originally present in the *glorlc?* GCTs. An *id* has been reconstructed a posteriori taking it from the first non-null member table (e.g. for the south, if there is a BCS counterpart, the *id* comes from *SFSbcs*, if there is none it goes on to DECcam, VISTA etc. however if the DECcam counterpart is not the one in the master associations made by SF for such *id* the value derived from BCS is preserved). This information was added as a last step and has not been used in the comparisons described here, and as said does not guarantee that a counterpart set in the "LC" GCTs is exactly the same in the "SF" GCTs (one should always check the "unid").

Such information was used to define the photometric redshift tables *SFNzphot* and *SFSzphot* as additional GCT members, so that one can access z_{phot} for all entries which have it (those with a valid positive *id*).

All GCTs foresee since their origin a space for three more members: *usno*, *simbad* and *ned*, the so-called "external tables" available since ever in the Milan DB in

correlation with the X-ray tables. Pointers to USNO, SIMBAD and NED counterparts have been added to all GCTs in the same way, which is *unlike* the incremental procedure used in the past for XLSS (the same used to build *glorlcn* and *glorlcs*, which introduced also "standalone external counterpart sets": these do not exist in the current approach).

The new procedure, supported by scripts *insertnorthexternal* and *insertsouthexternal*, to be run separately on each external catalogue, takes the counterpart sets of all X-ray sources which have a known correlation with an external catalogue, and computes simultaneously the distance with all other SF photometric members which are within the canonical distance of 0.7'' or 2''. Otherwise distance is set to undefined (99). Then it computes a best distance from counterpart to external catalogue (the smallest of those defined). It considers only the cases with a defined best distance, and solves the possible ambiguities when a counterpart sets is associated to more external objects (rare but possible) taking the closest one.

This may exceptionally result in more than one counterpart sets pointing to the same external objects (this usually occurs when the external object is associated to an IRAC, GALEX or WISE counterpart present in more than one counterpart set due to the larger positional uncertainty). This usually excludes from association the second farthest (and next farthest) external objects (sometimes they are rightfully spurious or redundant), but this is felt not a problem since they can always be known from the correlation tables and via the tool described in section 8.

6.2. Hunting ambiguities

The above checks have shown the presence of some ambiguities (more or less physiological and unavoidable) i.e. the same individual band detection could be attributed to more than one entries in the panchromatic table, or the same object in the panchromatic table could be attributed to more than one counterpart sets in the GCT.

I decided therefore to make an automatic screening of there two cases: the intrinsic ambiguities in the SF tables are described in 6.2.1; those in the GCTs in 6.2.2; possible solutions in 6.2.3.

6.2.1. Intrinsic panchro ambiguities

The 14 photometric SFN and SFS panchromatic tables contain from 1 (*idK* for *SFNwircam*) to 6 (*idu*, *idg*, *idr*, *idi*, *idy*, *idz* for *SFNcfht*) identifiers for the various band detections. These identifiers derive from the original master associations by SF.

Ideally one given band *id* shall occur only once in a table, and also a given combination of band *id*'s shall occur only once. Instead I found that multiple appearances

are possible, so I did a systematic screening. In the remainder I call *mini unique identifiers* (for short *minid*) the concatenation of the band identifiers for a given object (e.g. 0:4650354:1338:0:0:0: for SFNcfht indicates a detection in g and r bands). One has the following cases:

- a) *fully redundant* cases: two (or more) entries with the same minid (e.g. northern VISTA zK 309265:0:0:0:592081:)
- b) *partially included* cases: the minid of one entry has a few non-null band id's like northern VISTA YHK 0:414193:0:1429389:381502: which do match another object in those bands like 83504:414193:439810:1429389:381502: while the other bands are null in one entry, and not null in the other.
- c) *missed merger* cases: the minid's match in some bands, while the other bands are reciprocally null in one entry and not null in the other. Compare again from northern VISTA zY 161748:475402:475697:1415442:0: detected in zYJH, and 161748:475402:0:0:434839: detected in zYK, which could be merged in a single zYJHK entry
- d) *intrinsically ambiguous* cases: the minid's match in some bands, but the other bands are different detections. Compare northern VISTA Y 0:203561:0:0:176849: 696532:203561:202069:1143763:177023

Multiple appearances occurs in all surveys and bands with the exception of SDSS (all bands), CFHT u and BCS g. They are usually just double occurrences for optical/NIR tables but can reach multiplicity 5 for the IR and UV tables.

Multiple appearances occurs *only* for objects in the panchromatic tables which derive from the original association tables by SF. They *never* occur for entries flagged "appended" (see 4).

CFHT shows a limited number of double appearances (57, 77, 21, 64 respectively for the griz bands) affecting 221 distinct *survey* id's (or my seq), all corresponding to distinct minid's. The vast majority are ordinary couples which show the same identification in some bands (in the maximal case griyz) and a different one in other bands (e.g. u). Compare i.e. the two entries with id 1711391 and 2873647. Or there are cases like same gri, different u, one counterpart set has yz detection and the other null detection in yz, or a wide variety of combination. The only peculiar one is a "crossed" case where an r-band flagged duplicate 0:4650354:1338:0:0:0: points to an ri-band ambiguous 0:96860:1338:3711922:0:0:, which in turn points to an i-band case 0:4650575:0:3711922:0:0:.

Northern SFNvista has similarly a limited number of double appearances (22, 22, 27, 25, 31 respectively for zYJHK), affecting 120 distinct *survey* id's, but only 92 distinct minid's ! The maximal detections are 7 couples with the same zYJHK identifiers identical (fully redun-

dant) which ideally should be "collapsed" in 7 objects. The detections with matches in one, two or three bands are a mixed bag of 19 fully redundant, 8 partially included, 4 missed mergers and 8 really ambiguous cases.

Southern SFNvista is alike (35, 38, 34 doubles respectively in JHK, 109 minid's distinct out of 149, 10 redundant couples of maximal JHK, mixed bag with a prevalence of redundant cases).

DECAM has a larger number of double appearances (79 g, 86 r, 173 i, 356 z) including even a triple in the iz bands. Only 689 minid's out of 758 id's are distinct. Maximal detections (griz) are 22 fully redundant couples. The matches in one, two or three bands are the usual mixed bag (not unlike VISTA).

BCS has a small number of double appearances (24, 61, 31 in riz, none in g !). All minid's are distinct, the 3 maximal couples (griz) match in riz and differ in g so they should be left alone. The cases with matches in one or two bands are usually ambiguous with a different detection usually in g (or some other band), so not unlike CFHT.

UKIDSS has a small number of doubles (18, 3, 22 in JHK), with 28 distinct minid's out of 48 id's. Just two redundant couples are maximal detections in JHK. All the one or two band matches are fully redundant but two partially included and one ambiguous.

WIRCam has only the K band, so it presents 21 couples which are fully redundant (same id in the K band).

The first of the "2''-surveys", GALEX, shows a much larger number of multiples (2127 NUV and 674 FUV in the north, 499 and 199 in the south, with multiplicities up to 5). The distinct minid's are 2455 out of 4858 id's in the north (558 out of 1126 in the south). The maximal (NUV and FUV) redundant detections are 538 couples (N) and 152 (S). Considered that there are just two bands, one can have only fully redundant cases or partially included ones. Or, for multiplicities above 2, combinations thereof. For instance there are two entries (id 8 and 95257) with redundant minid (0:102823: which are both included in 1288382:102823:.

The other two-band 2''-survey, IRAC, shows similarly a larger number of multiples (5317 3.6 μ m and 4758 4.5 μ m in the north, 3764 and 3544 in the south). 7771 distinct minid's in the north (4908 in the south). 4080 and 3221 maximal redundant detections. So not unlike GALEX.

The remaining 2''-survey, WISE, has somewhat less multiples (144, 964, 421, 171 in W1/2/3/4 for the north, 816, 697, 224, 92 for the south). The distinct minid's are 2698 or 1618. The maximal detections (in all four bands) are very few (43 cases in the north, just 8 in the south). In the north they are not only double, there is one triple redundant, and 5 single cases. Of them 3 are in partial include relationship with two W1/2/3 or one W1/2 case. One relates to a W1/3/4 ambiguous (W2 is different). The remaining one is a complex case: 531062:0:0:358582 and 531062:478091:511650:0: are in "missed merger" relationship between them, and both shall be included in

531062:478091:511650:358582:. Similar mixed combinations occur also in the cases for one, two or three bands.

6.2.2. GCT ambiguities

The 2×2 GCTs contain 8 (N) or 6 (S) pointers to their non-X-ray member tables. Ideally each counterpart set (record in the GCT) shall contain a unique combination of pointers (seq into the SF? panchromatic tables), some of which can be null. Each given pointer shall appear only once (each counterpart shall appear in a single counterpart set), and each combination of pointers ("unid" as defined in 6.1.3) shall appear in a single counterpart set.

For the `glorsfn` and `glorsfs` there are no ambiguities in this respect. Each unid (and associated id) appears only once if one limits to the non-X-ray members. Of course this does not prevent the ambiguities described in 6.2.1 (if an id is ambiguous within the panchromatic tables). If one extends the check to the X-ray members one can have another form of (apparent) ambiguities, i.e. two X-ray sources (exceptionally three like the case, in the south, of `Xseq` 224400, 224401, 224467 and 202764, 214634, 214635) which share a candidate counterpart set. But this fact is pretty normal, and involves 588 cases in the north and 550 in the south (the affected X-ray sources may be band-merged ambiguities, flagged as "suspect" or "divorced", but are most frequently just fully independent X-ray sources).

I remind that `glorsfn` and `glorsfs` correspond to the case where the association between optical, NIR, IR and UV is done first, and only later the candidate multiwave counterparts are associated with X-ray sources.

Instead `glorlcn` and `glorlcs` represent the case in which X-ray sources are first correlated directly with *each* optical, NIR, IR or UV survey, and only then these correlations are combined. This survey-level operation is analogue to the band-level merging done by SF in her procedure. In principle one could have the equivalent of the four cases listed in 6.2.1, but applied *at unid level* instead of *minid*. In practice (excluding the cases where the same counterpart set is associated to different X-ray sources, mentioned above as normal, and restricting to the check of the unid of the counterparts of *each* X-ray source with themselves) one finds that:

- a) there are *no* fully redundant cases (for the same X-ray source)
- b) there are *no* partially included cases
- c) there is a limited number (about 2% of the total number of counterparts) of missed mergers (2776 in the north and 1846 in the south)
- d) there is a larger number (between 12 and 8%) of intrinsic ambiguities (14880 in the north and 9401 in the south)

Note that missed merger and ambiguities may apply more than once to the same counterpart, and affect a rela-

tively small number of X-ray sources (e.g. 1832 for missed mergers in the north - 1319 in the south - and 4943 for ambiguities in the north - 3816 in the south). These cases often involve the surveys with larger positional uncertainty (so called 2" surveys, IRAC, GALEX, WISE).

As an example consider for south `Xseq=200014` the objects with unid `66786:0:27017:35869:0:0` and `0:62838:27017:35869:0:0` which associate respectively one BCS and one DECam object (at 0.74") to the same VISTA/IRAC combination. Ideally they could be merged in a single counterpart `66786:62838:27017:35869:0:0`. Note that `glorsfs` actually contains already such a counterpart, so the procedure by SF was already correctly associating BCS 66786 with DECam 62838, unlike `glorlcs`. But `SFSdecam.seq=62838` has an `sflags` of "supplemented" in g. So obviously it is just the g band position which is 0.04" too far !

Note also that because of the nature of the LC procedure (correlating X-rays with each other band first, and putting them together later) there are cases resembling missed mergers or includes which are fully normal "features". E.g. there is an object (south) `5440:0:0:0:0:0` (for `Xseq=200505`) which looks included in `5440:5110:4916:4708:0:2351`: but the latter refers to `Xseq=200467`. Actually BCS 5440 and its equivalent in the other surveys is the counterpart of `Xseq=200467` (and has even a Lidman spectrum !). For `Xseq=200505` BCS 5440 is rather farther out in the FoV, just within 10", while the positions in the other surveys are just beyond 10", and therefore ignored by the LC procedure.

It seems wise to ignore so far the ambiguities and defer fixing until some form of ranking of the counterparts has actually rejected most or all of the outermost ones.

6.2.3. Fixing ambiguities

Since we are able to spot and possibly flag the ambiguities, one may think to fix them with corrective actions.

Full redundancies in SF? panchro tables (two entries with different id and same minid) could be corrected dropping one of the entries. Or one could leave the SF? panchro table alone (for "historical" reference) and just repoint `glorsf?` to one of the entries.

Partial includes in SF? panchro tables (two entries of which one minid non null components are included in a "longer" minid) could as well be corrected dropping the shorter entry from the panchro table, or leaving it alone and repointing `glorsf?` to the longer entry.

Missed mergers in SF? panchro tables or in `glorlcs?` GCTs could be somewhat flagged, or even actually merged. This is easy for the GCTs which are "work in progress", while it should be handled carefully for the panchro tables (probably one should introduce new merged entries, preserve the old entries for the record and repoint the `glorsf?` GCTs to the merged entries).

For *intrinsic ambiguities* in SF? panchro tables or in glorlc? GCTs there is nothing which can be done but live with them and perhaps flag them.

A posteriori (after the ranking described below in 7.5) I looked again at the ambiguities to see how many of the more or less firm primary counterparts are affected, with the following, though not final, conclusions:

- the full redundant entries in the "0.7" panchro tables (see 6.2.1) are one or two handfuls per survey if one limits to the non-rejects
- they are definitely more for the "2" panchro tables (specially IRAC and somewhat GALEX)
- anyhow in all cases, of the two ambiguous entries with a different general id and the same band id's (minid) in one survey, they are associated to different objects in the other surveys, so at the end the effect is irrelevant (one will have a collection of magnitudes per band, and if, e.g., IR, NIR and UV are different objects with different magnitudes, while optical assigned the same magnitudes to nominally different objects, who cares?)
- for the GCT ambiguities (see 6.2.2) which affect only the glorlcn and glorlcs tables, one can narrow down the statistics. If one considers the ambiguities in which *both elements are primary counterparts* (in practice this means exclusively that they are both superrank=1, see 7.5.1), they are just 236 counterparts (all distinct X-ray sources) of which 66 missed mergers for the north (119 and 45 for the south). One has more cases (5588 in the north for 3043 distinct X-ray sources, 3753 for 2284 in the south) where one of the elements is a primary counterpart and the other a secondary or rarely a reject (of 5588 733 are missed mergers in the north, respectively of 3753 594 in the south),

So I decided to do nothing so far for panchro table ambiguities, while for GCT ambiguities I store offline a list of the offending cases for later inspection.

6.3. Making views

The names of the VIEWS associated to the GCTs described in sections 6.1.1 and 6.1.2 are respectively NEWSFN, NEWSFN, NEWLCN and NEWLCS. They are released in the category "XXL merged tables (views)" soon after the main XXLN and XXLS catalogues, and share a common subset of columns with the same name.

- Only the X-ray source identifier Xseq, plus the IAU catalog name, the soft and hard band identifier and the X-ray coordinates are taken from the corresponding XXLN or XXLS catalogue,
- There is a family of *pointers* with names like SFsurvey which point into the panchromatic tables. Note that the name has no N/S designation (so SFwise may point into SFNwise or SFSwise depending on the fact

the view is a NEW*N or NEW*S one, while e.g. SFcfht obviously points into SFNcfht and SFbcs points into SFSbcs.

- Another family of *pointers* point into the external tables for USNO, SIMBAD and NED.
- No further columns from the member tables are included, they shall be accessed by name as said below.
- There are instead the P and LR ranks and subranks described in 7.5 at pag. 37, the corresponding best value of P and LR Pbestall and LRbestall (see 7.4), and the superrank (see pag. 40 in 7.5.1) and comparison (see pag. 41 in 7.5.2).
- No further columns from the GCTs are included, but they can be accessed by name, knowing the name and prefixing it with the GCT name (for instance the g-band CFHT probability for NEWLCN can be accessed as glorlcn.Pcfhtu).

Note that currently there are no correlation tables between these views and the other tables, since the photometric panchro tables (see 4.1) and the external tables (see 6.1.4) are defined as *member tables* of the GCT. Therefore one can access items in such tables, like magnitudes or photometric redshift, just calling them by name using the trick described in section 9.

7. Ranking the associations

The published works done for XMM-LSS and XXL so far (as well as the preceding reports of this series) up to Chiappetti et al. (2013) have ranked associations according to the chance probability after Downes et al. (1986). The XXL-GAMA Matching Group has first used (Bongiorno & Brusa, 2014) the Likelihood Ratio (LR) method originally proposed by Sutherland & Saunders (1992) and used e.g. by Brusa et al. (2007).

Here I attempt to use and compare both, using a direct implementation in mysql within the database.

7.1. Methodology and notation

The chance probability after Downes et al. (1986) is given by the simple formula

$$p = 1 - \exp(-\pi n(< m) d^2) \quad (1)$$

where d is the distance between the X-ray source and the candidate counterpart, and $n(< m)$ is the density (per square arcsec if d is in arcsec) of objects brighter than magnitude m in the whole survey.

Such density is computed from the number count of objects brighter than m , $N(< m)$ and the area A covered by the survey

$$n(< m) = N(< m)/A \quad (2)$$

Concerning the method by Sutherland & Saunders (1992) one has the following notation. One starts from the number count of objects at magnitude m (e.g. using

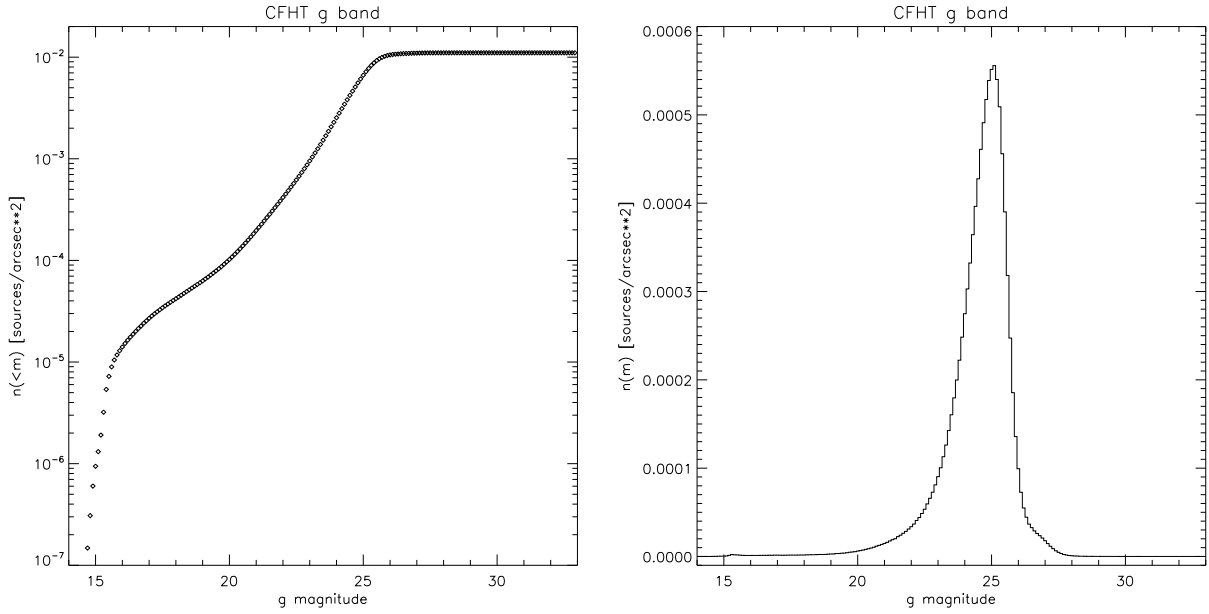


Fig. 10. The cumulative density $n(< m)$ and the density $n(m)$ for CFHT g sources.

an appropriate binning, like $\Delta m = 0.1$), $N(m)$, and computes a density in the whole survey.

$$n(m) = N(m)/A \quad (3)$$

Analogously one computes as function of magnitudes the number count $Q(m)$ of putative true counterparts, taken as unique objects within a suitable radius, which is assumed here as $r = 3''$. Assuming there are N_i such counterparts, the covered area is given by

$$A_i = \pi r^2 N_i = \pi 9 N_i \quad (4)$$

and one can compute a density

$$g(m) = Q(m)/A_i \quad (5)$$

One can then compute an expected density

$$q(m) = g(m) - K_i n(m) \quad (6)$$

where $K_i = N_i/N$ and N is the total number of (non-unique) objects within the radius r . This should ensure that $q(m)$ does not go negative. Finally one can compute the likelihood ratio LR as

$$LR = (q(m) \exp(-0.5d^2)/2\pi) / K n(m) \quad (7)$$

$$LR = (\exp(-0.5d^2)/2\pi)(g(m)/K n(m)) - 1 \quad (8)$$

$$LR = (\exp(-0.5d^2)/2\pi) f(m) \quad (9)$$

which can be separated in a term depending on the X-ray to optical distance and one, $f(m)$, depending purely on magnitude.

7.2. Implementation in mysql

Of the elements entering equations 1 or 7, the terms $n(< m)$ and $f(m)$ depends only on the magnitude of the potential counterpart, while the rest of the equation depends on the X-ray to counterpart distance.

Therefore (for each survey and band) one can compute a chance probability and an LR for each counterpart set in a GCT computing on the spot the X-ray to counterpart distance d , and looking up the magnitude dependent term in a previously generated tabulation. This approach is different from the one I used in the past for probability, where I fitted $n(< m)$ to a straight line in log log space, and used the coefficients of the fit to estimate $n(< m)$ at the counterpart magnitude.

Using a precomputed tabulation has the advantage one can easily handle all possible surveys and bands (while earlier I did the fit only for a few selected filters), and can do all computations within mysql, i.e. within the database (while previously I did the fits in IDL).

7.2.1. The densities

As explained above in 4, during ingestion I saved a "band density table" (*bdt*) for *all* objects in a filter. This table contains the truncated coordinates *r1,d1* and the magnitude.

I can compute the area A covered by the survey in a particular band from the number of filled pixels of 0.01×0.01 deg with the statement

```
select count(distinct r1,d1)/10000 from bdt
```

The count densities $N(< m)$ and $N(m)$ for the entire survey can be computed making an histogram with a mag-

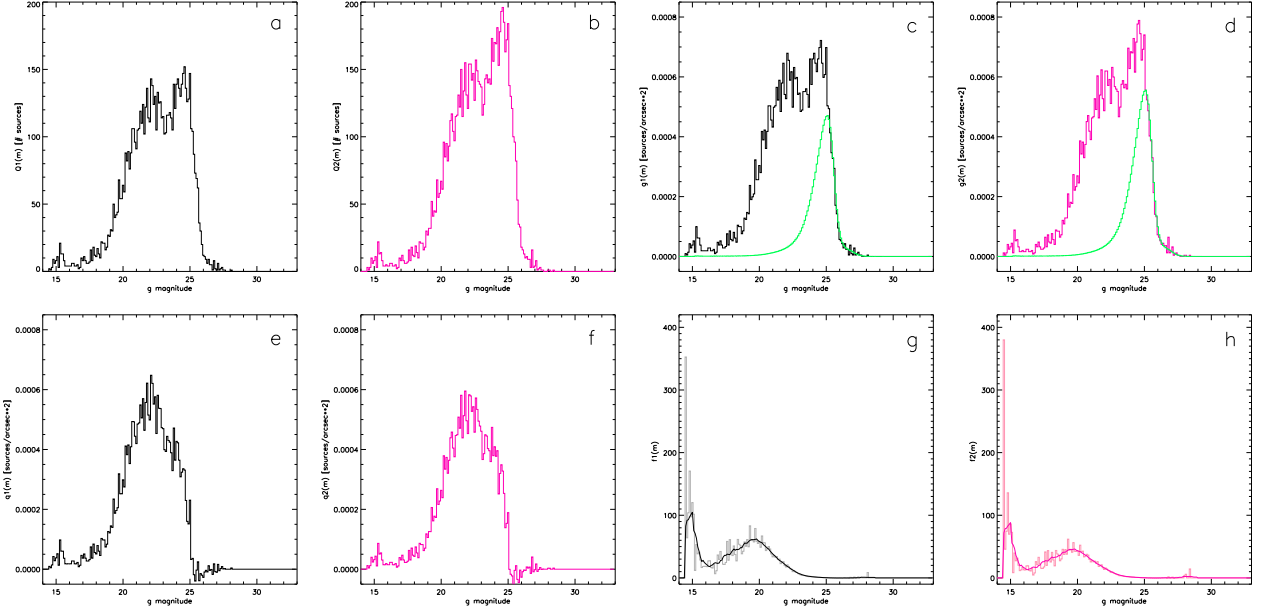


Fig. 11. Steps for computation of LR. Black (gray) curves are for the unique candidate case. Magenta (pink) curves are for the closest candidate case. Panels a/b show the number count $Q_1(m)$ or $Q_2(m)$, whose different height reflects the different total count $N_1 < N_2$. Panels c/d show $g_1(m)$ and $g_2(m)$ which are normalized by area. The green curves are respectively $K_1 n(m)$ and $K_2 n(m)$ which are subtracted to give $q_1(m)$ and $q_2(m)$ reported in panels e/f. Panels g/h report the raw and smoothed values of the final magnitude dependent terms $f_1(m)$ and $f_2(m)$ used for the computation of LR

nitide bin $\text{magstep} = \Delta m = 0.1$ from the band density table, e.g. for the i -th bin at magnitude mag_i

```
select count(*) from bdt where mag between 0 and magi
```

```
select count(*) from bdt where mag between magi and magi+magstep
```

I combined the loop for all magnitude bins between the brightest and faintest magnitude, the area computation, and equations 2 and 3 into a mysql stored procedure so I can e.g.

```
call dens('SFNcfht_g_dens',0.1)
```

which reads the band density table for CFHT band g `SFNcfht_g_dens`, and computes the histograms of $N(m)$, $N(< m)$, $n(m)$, $n(< m)$ for a binsize of 0.1 mag and stores them in a small "band density summary table" `SFNcfht_g_ndens` once forever.

7.2.2. The auxiliary tabulations

Ideally, I'd want an equivalent stored procedure to compute the terms for LR. There are some complications due to the fact dynamic SQL statements (i.e. involving a table or column name as argument of the procedure) are inefficient. So one should first prepare a temporary table *with a fixed name* from the join of the GCT, the SF photometric table and the correlation between the latter and the X-ray

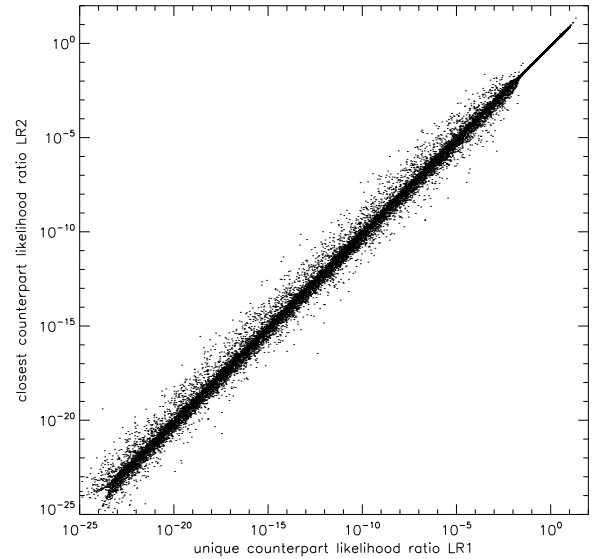


Fig. 12. Relationship between the likelihood ratio computed for the unique candidate case (LR_1) vs the closest candidate case (LR_2) for CFHT band g and GCT `glor1cn`.

table, which extracts the magnitude in the wished band for the objects within $3''$.

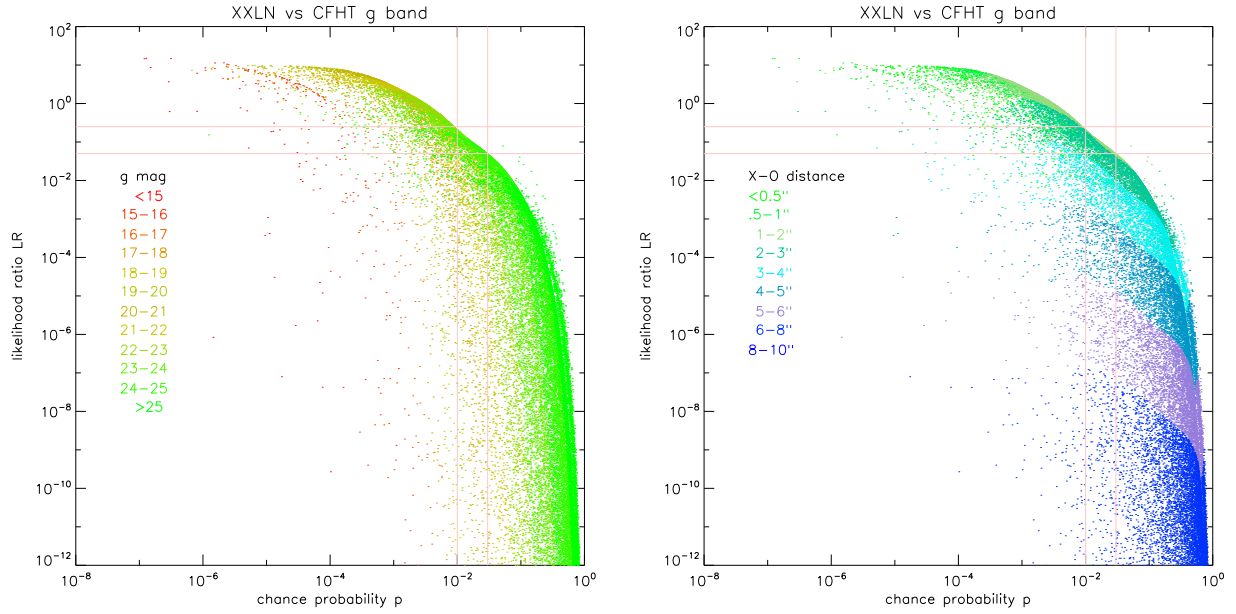


Fig. 13. Likelihood ratio vs chance probability. The two panels plot the same data, but the colour code is for magnitude in the left hand panel and for X-ray to optical distance in the right hand panel. The vertical fiducial lines are for $p = 0.01$ and $p = 0.03$, the horizontal ones for $LR = 0.25$ and $LR = 0.05$.

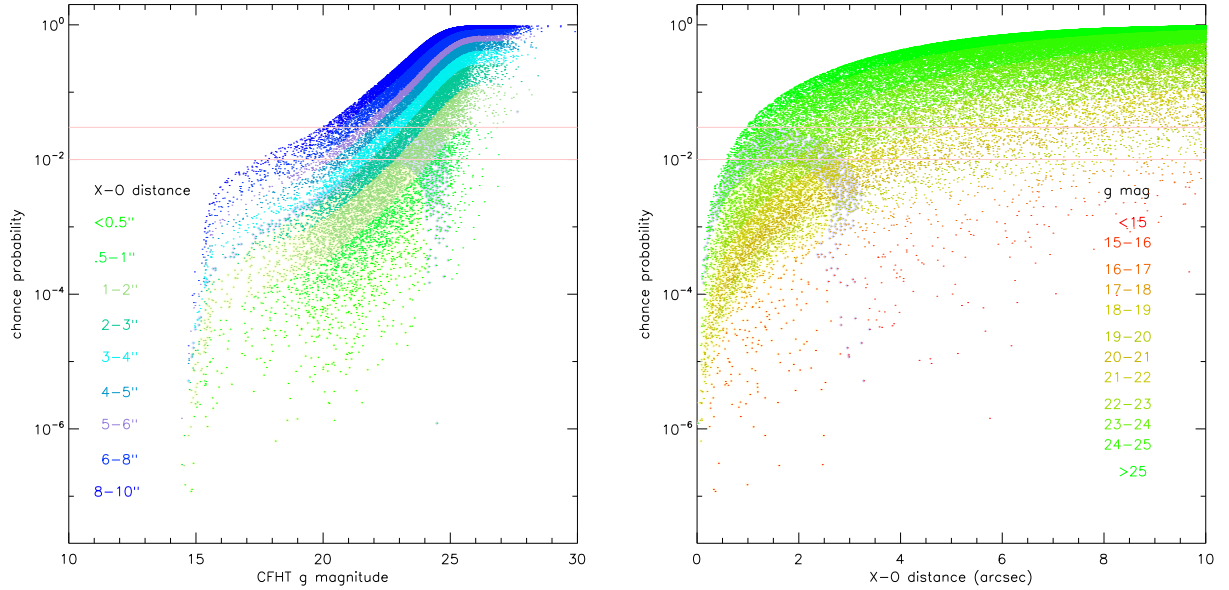


Fig. 14. Chance probability vs magnitude (left hand panel) and vs X-ray to optical distance (right hand panel). Each panel uses a colour coding as in Fig. 13 for the other parameter (distance in the left hand panel and magnitude in the right hand one). Fiducial lines are for $p = 0.01$ and $p = 0.03$. The points having an LR above 0.25 are surrounded by a yellow diamond, and those with LR between 0.05 and 0.25 with a gray-blue background.

Then one can call a stored procedure which does all the computations in equations 4 to 7, e.g.

```
call densq('SFNcfht_g_ndens',0.1)
```

which reads the band density summary table for CFHT band g created by the previous step and computes

$Q(m), g(m), q(m), f(m)$ for a binsize of 0.1 mag and stores them in a further "band q table" SFNcfht_g_qdens using the same magnitude interval in the band density summary table.

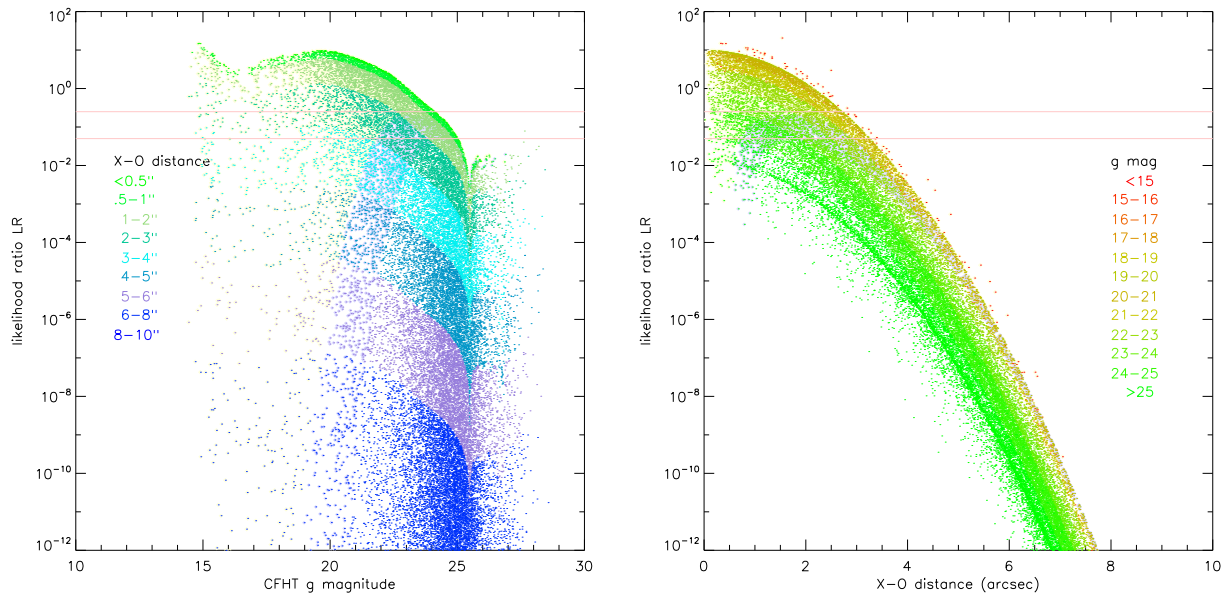


Fig. 15. Likelihood ratio vs magnitude (left hand panel) and vs X-ray to optical distance (right hand panel). Each panel uses a colour coding as in Fig. 13 for the other parameter (distance in the left hand panel and magnitude in the right hand one). Fiducial lines are for $LR = 0.05$ and $LR = 0.25$. The points having p below 0.01 are surrounded by a yellow diamond, and those with p between 0.01 and 0.03 with a gray-blue background.

I introduced two additional features in the stored procedure. One is that I do the computations for two different cases (so I have two sets of parameters):

- in the first case the count $Q(m)$ is on N_1 objects, i.e. the counterparts within $3''$ which are *intrinsically unique*. So I take only the cases where an X-ray source has exactly one counterpart.
- in the second case the count $Q(m)$ is on N_2 objects, i.e. I take the closest counterpart within $3''$ if there is more than one.

The second feature is that the final parameter $f(m)$ (actually two sets $f_1(m)$ and $f_2(m)$) is smoothed with a running average of ± 5 bins. While this is unnecessary for $n(< m)$ or $n(m)$, the data in the $3''$ domain will otherwise be too noisy.

7.2.3. The lookup

Since one wants to add the probability p and the likelihood ratio LR for all counterparts in a GCT, one has to apply equations 1 or 7, and to interpolate respectively $n(< m)$ or $f(m)$ from the band density summary table or the band q table.

I wrote a stored function `lookup(mag)` to do the linear interpolation in the nearest bin of the precomputed tabulations, so I could write the formulae as statements (assuming $d = \text{dist}(x.ra_corr, x.dec_corr, o.ra, o.dec) * 3600$)

```
set prob = 1.-exp(-pi()*lookup(mag)*d*d)
set lr1 = lookup(mag)*exp(-0.5*d*d)/2/pi()
```

however this faced two technical issues:

- One is that stored functions are even more picky than stored procedures about dynamic SQL. So I have to rely on the fact `lookup` works on a fixed name temporary table, which I have to prepare in advance containing the magnitude and $n(< m)$, or the magnitude and the smoothed $f(m)$. In practice this means I have to work in series, adding p then LR for each survey and band for all counterpart sets in the GCT, rather than in parallel (all p s and LR s for all surveys/bands for each counterpart set).
- The second is that despite the fact the summary table and q table are very small (less than 200 bins), the overhead involved in a call to `lookup` is not negligible. Unless appropriate measures are taken, a single call takes 0.05 s, which means the computation of one parameter for one band for some 50000 counterparts in a GCT takes 38 minutes ! Multiply for 48 bands and 2-3 parameters (p , LR_1 , LR_2) and this is clearly too long. A simple solution involves adding an SQL index to the summary table or q table (despite their small size). This cuts the speed by a factor 736, and now the computation for a single parameter and an entire GCT goes down to 3 s !

7.3. Closer view to a case

As example, I show the case of CFHT g band. The densities $n(< m)$ and $n(m)$ are reported in Fig. 10.

The various steps for the computation of LR are shown in the panels of Fig. 11. The odd panels (a,c,e,g) show the computation for intrinsically unique candidates, while the even panels (magenta curves) show it for the closest candidates. I get $N_1 = 7448$ unique candidates (since $N = 8789$, $K_1 = 0.85$) and $N_2 = 8789$ closest candidates ($K_2 = 1.0$ by construction).

I show also the empirical relationship between probability p and LR (actually LR_1) vs magnitude and distance for all candidate counterparts (with a defined g band measurement) in GCT `glor1cn`.

LR_1 and LR_2 correlate very well on their entire range (although the useful range will be only the one for $LR > LR_{th}$, as explained below), however with LR_2 being a factor 1.3 lower than LR_1 , as shown in Fig. 12. Since LR will be used mainly for ranking counterparts according the fact LR is higher or lower, it is possibly indifferent to choose one of the two ways, and the difference can be regarded as an arbitrary normalization.

The two panels of Fig. 13 show the *same* plot of LR vs probability, where the colour code from red to green is for 1-mag magnitude ranges, and the colour code from green to blue is for increasing X-ray to optical distance.

I plot some fiducial lines. Traditionally so far, with the chance probability used for ranking like in Chiappetti et al. (2013), we consider the cases with $p < 0.01$ as *good* and those with $0.03 > p > 0.01$ as *fair*. The likelihood ratio analysis usually takes objects with $LR > LR_{th}$, where Bongiorno & Brusa (2014) takes the threshold $LR_{th} = 0.25$ (Brusa et al. (2007) uses 0.40). By mere chance the $p = 0.01$ and $LR = 0.25$ lines intersect exactly where they cross the envelope of the (LR, p) plot. For analogy I introduced a lower more conservative threshold $LR = 0.05$ in correspondance of the intersection with $p = 0.03$.

In Fig. 13 the points *above and to the left* of the fiducial ones are the better candidates, while those below and to the right shall be rejected.

Fig. 14 shows the dependency of chance probability on magnitude and distance (the better objects are those *below* the fiducial lines). For instance a very far object (blue point in the left hand panel) can be a good candidate only if brighter than 17. Similarly a very faint object (green in the right hand panel) can be an acceptable candidate only if the distance is less than $1''$. The yellow and gray backgrounds mark the objects which are also good or fair according to the LR thresholds. One can see that LR is more restrictive, and tend to favour closer objects (not surprising since it has a $3''$ radius built in).

Fig. 15 shows the dependency of LR on magnitude and distance (here the better objects are those *above* the fiducial lines). The yellow and gray backgrounds mark here the objects which are also good or fair in probability, and once again the probability criterion is more relaxed accepting some more distant or fainter objects.

7.4. Adding ranks to GCTs

After the generation of the band density summary tables (per 7.2.1) and of the "q tables" (per 7.2.2), I could proceed to the computation of the chance probability P and of the likelihood ratio LR_1 for all counterpart sets in all GCTs for each survey and filter. The intermediate parameters used for the computation are reported in Figures 16 to 27.

Given that both probabilities and likelihood ratios (after visual inspection) are commensurable across surveys and bands, one can now pass to some form of ranking.

A trivial ranking is *order ranking*. Simply assign a position rank (1 to n) for all n counterpart sets of a given X-ray source by increasing P (the lowest the best), or decreasing LR (the highest the best). The best counterpart set has rank 1, the next has rank 2 and so on. So one can have up to 28 couples of ranks for the northern GCTs (one per band per survey).

Another possibility is *threshold scoring*. Classify a counterpart as good (score=1), fair (score=2) or bad (score=3) based on the thresholds discussed above ($P < 0.01$, $0.01 < P < 0.03$, $P > 0.03$ or $LR > 0.25$, $0.05 < LR < 0.25$, $LR < 0.05$). Here for each X-ray source one can in principle have one or more or no counterpart scored good, and so on. Again one can have up to 28 couples of scores per band per survey.

One can then group ranks and scores per survey in two ways. One can take an *average* (rank or score) or a *best* rank or score. The best is based on the best (lowest) probability or best (highest) LR for the bands in the survey which have a defined P or LR. Then one ranks or scores such best P or best LR. The average instead simply averages the defined P or LR in all defined bands. Conventional average ranks or scores can be fractional. For instance if a source ranks first (1) in the u and g band, and second (2) in the i band, and is not detected in the other bands, the average rank will be 1.33 (one has to divide by the number of used bands).

This allows for instance for the northern GCTs to have best P and LR, and 4 sets of indicators (average or best, rank or score) for each of the 8 surveys. De facto one considers only 7 surveys, as the values for WIRcam coincide with the ones for the K_s band (only band) of WIRcam.

One can extend the same reasoning to broader categories, and computing best P and LR and the indicators (average or best, rank or score) for optical (combining CFHT and SDSS bands for the north, BCS and DECam for the south), NIR (combining VISTA, UKIDSS and WIRcam for the north), or IR (combining IRAC and WISE). There is no true UV set (it coincides with the GALEX one), nor a true NIR for the south (it coincides with VISTA).

Finally one can extend the reasoning to all bands, and compute best P and LR for all bands, and their 4 indicators (average or best, rank or score) also for all bands.

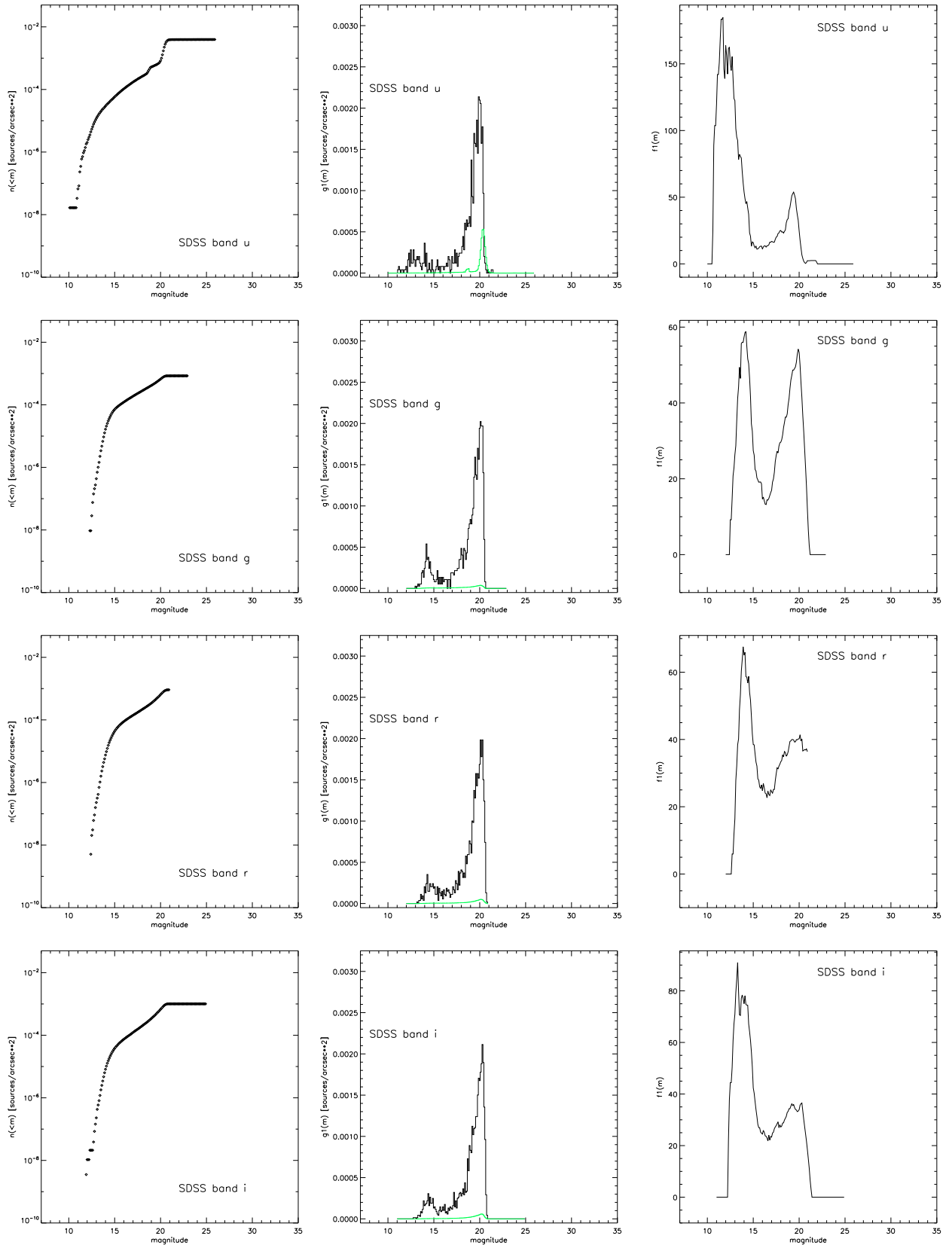


Fig. 16. For each band the leftmost column plots the density $n(> m)$ used for probability computation, the central column the parameter $g_1(m)$ and the scaled $K_1 n(m)$ (in green), and the rightmost column the smoothed $f_1(m)$ entering the computation of LR. All panels but the ones in the right hand column have the same scale. This page reports the plots for SDSS ugr bands.

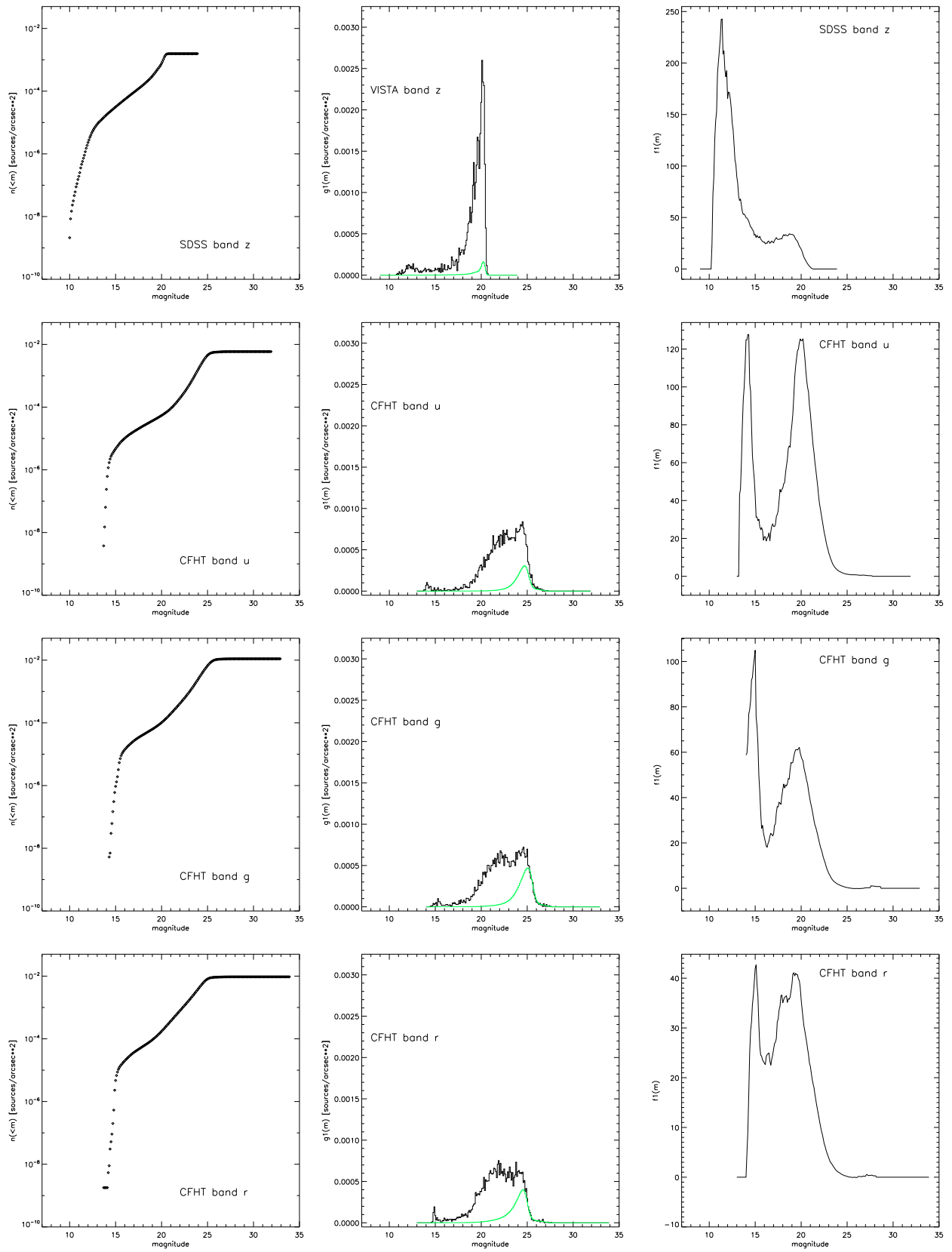


Fig. 17. Continuation of Fig. 16 for SDSS z, and CFHT ugr

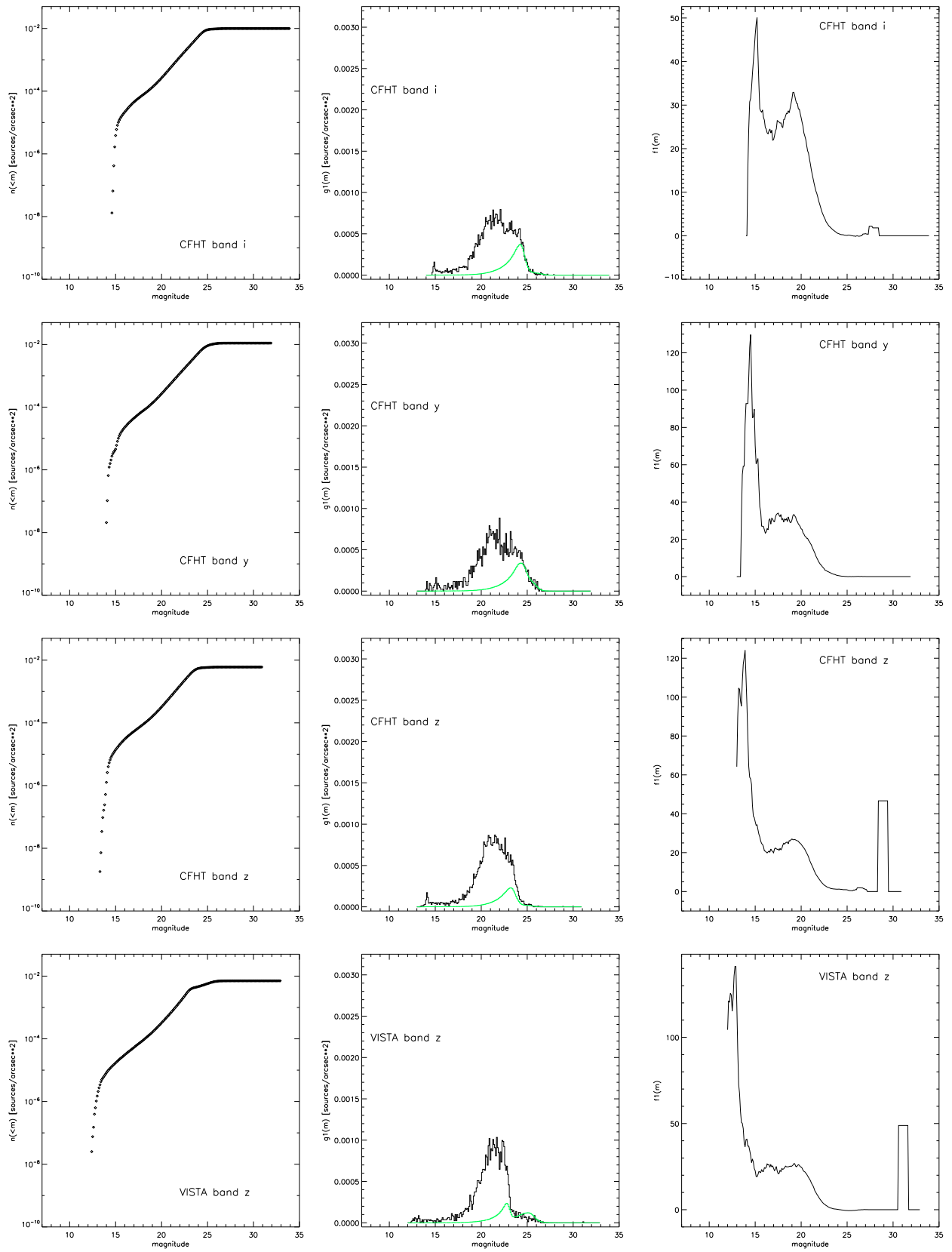


Fig. 18. Continuation of Fig. 16 for CFHT iyz and (north) VISTA z

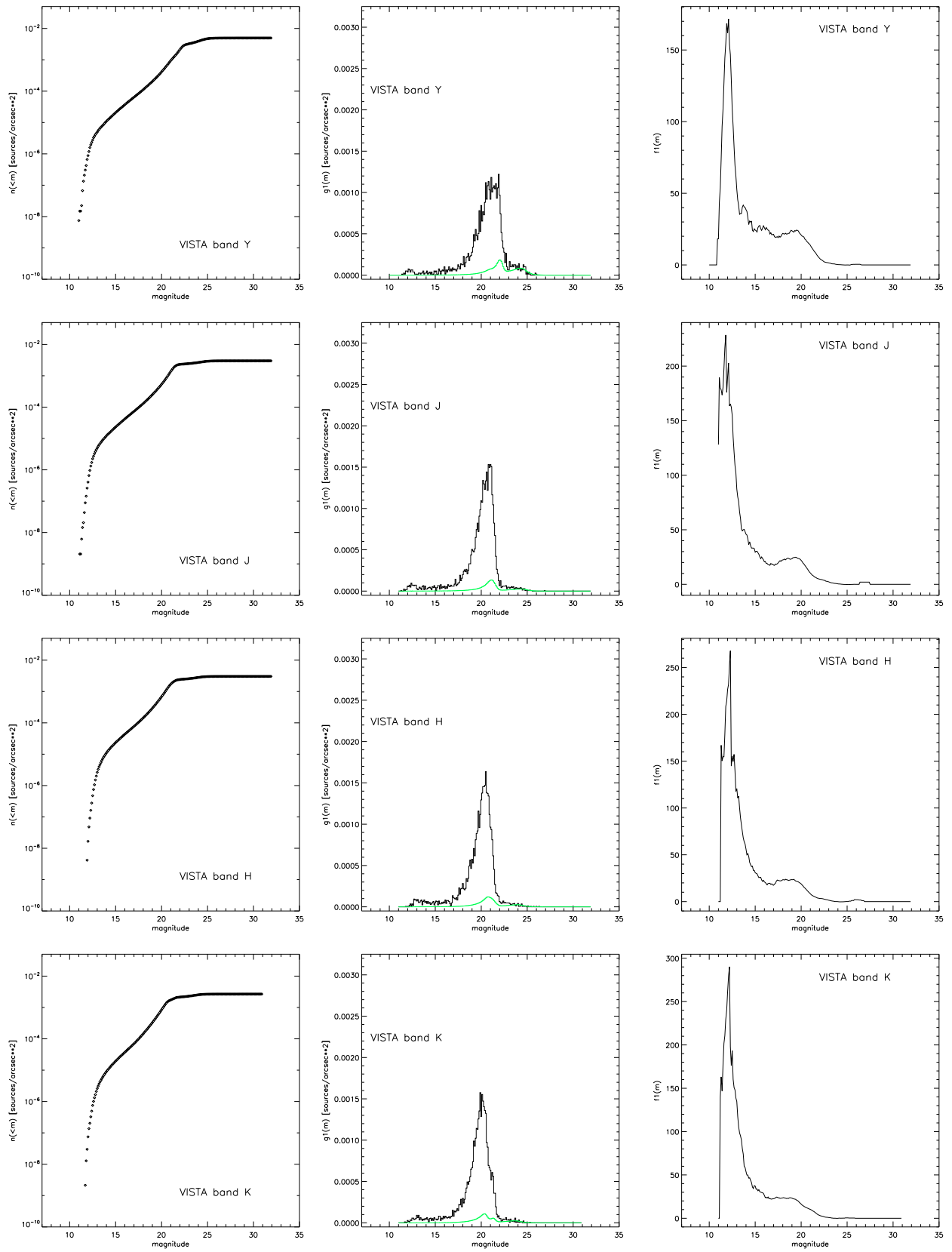


Fig. 19. Continuation of Fig. 16 for (north) VISTA YJHK

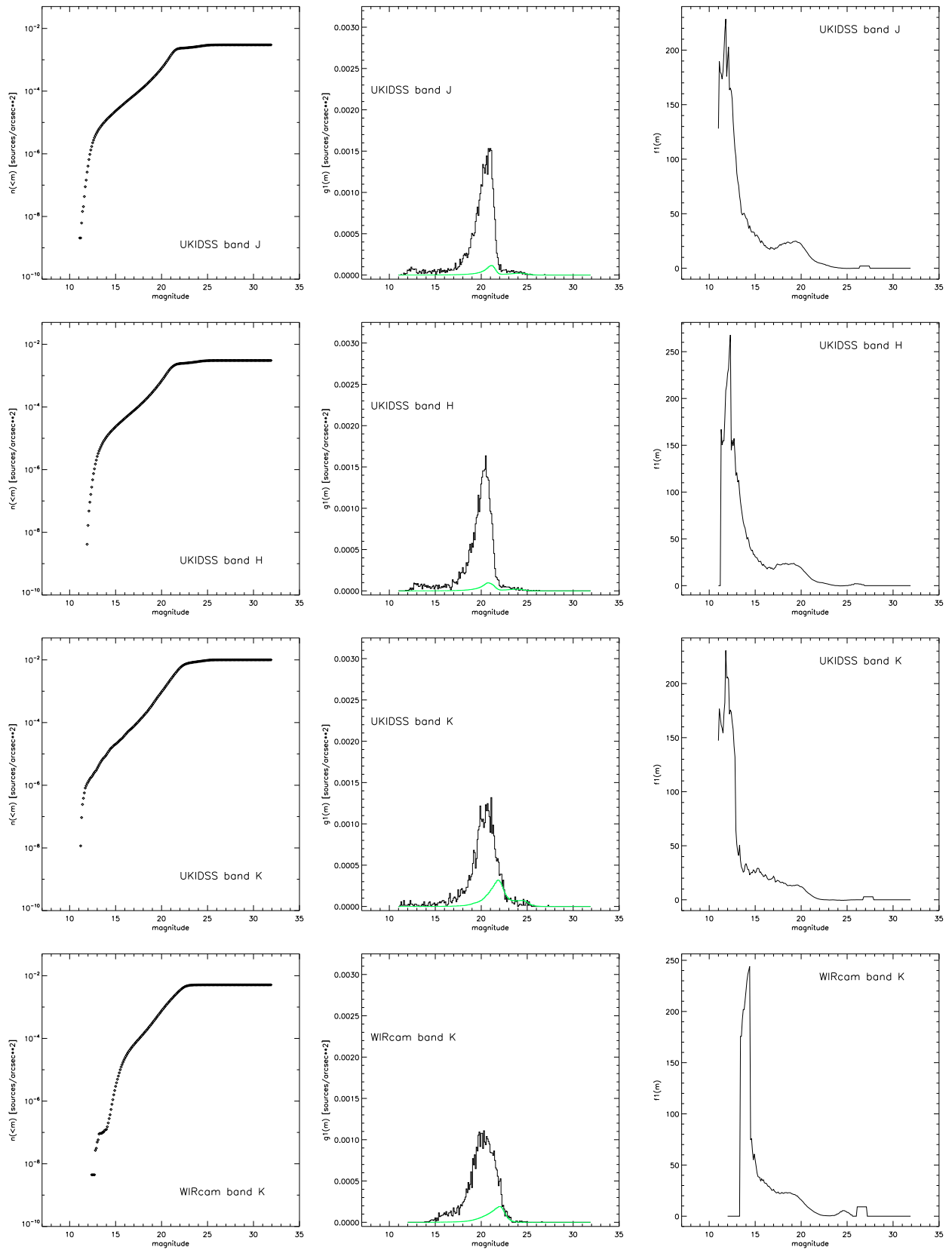


Fig. 20. Continuation of Fig. 16 for UKIDSS and WIRcam bands

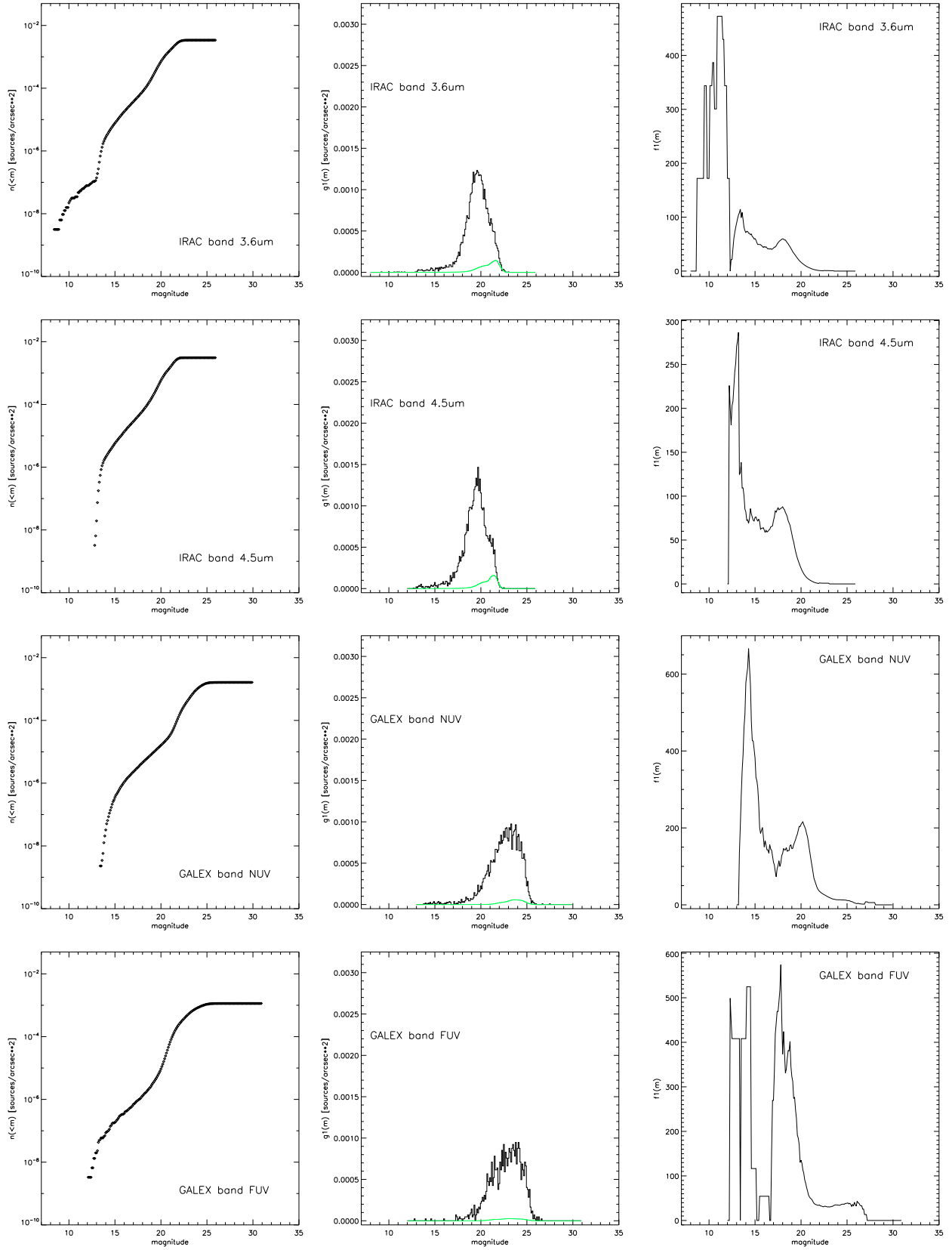


Fig. 21. Continuation of Fig. 16 for north IRAC and GALEX bands

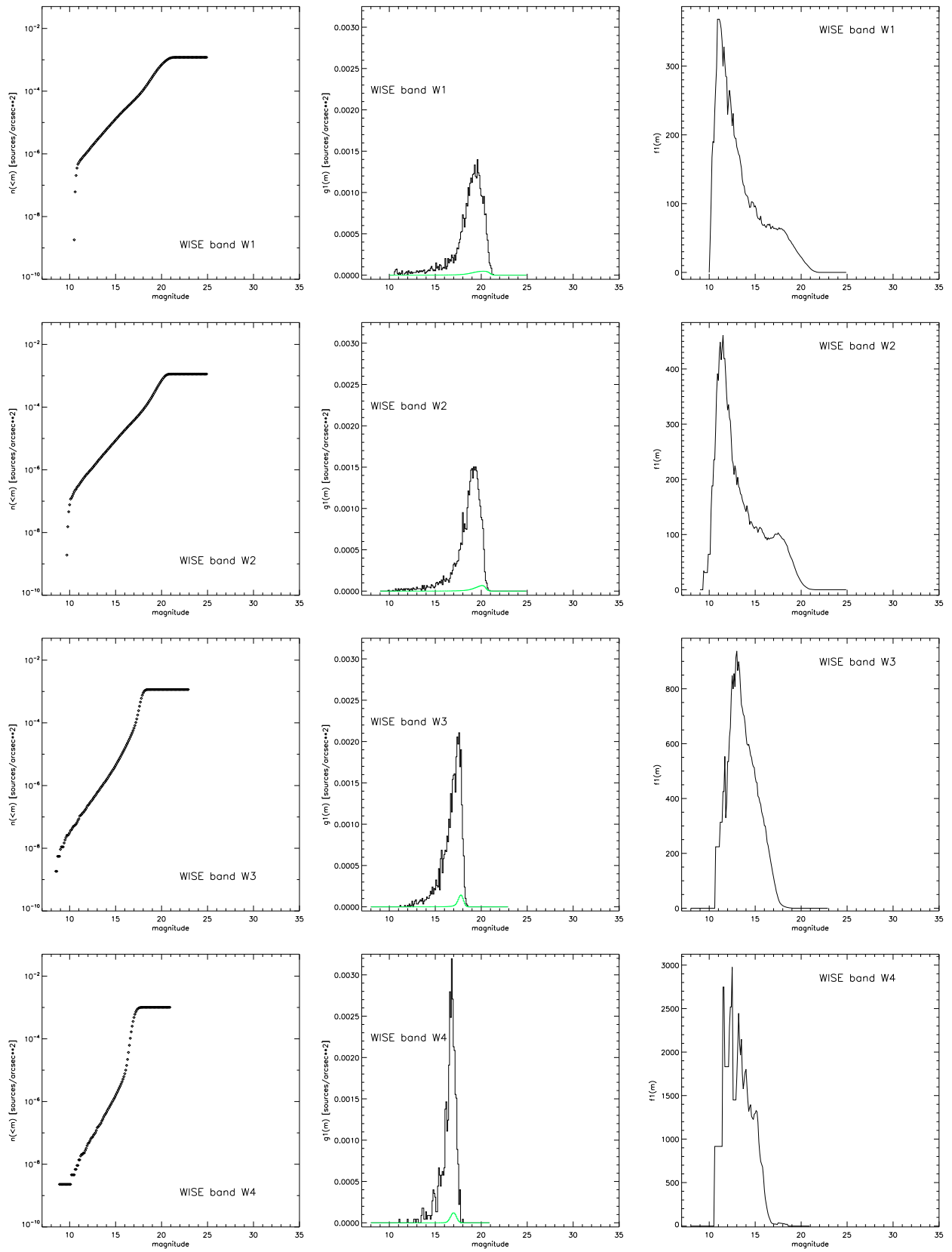


Fig. 22. Continuation of Fig. 16 for north WISE bands

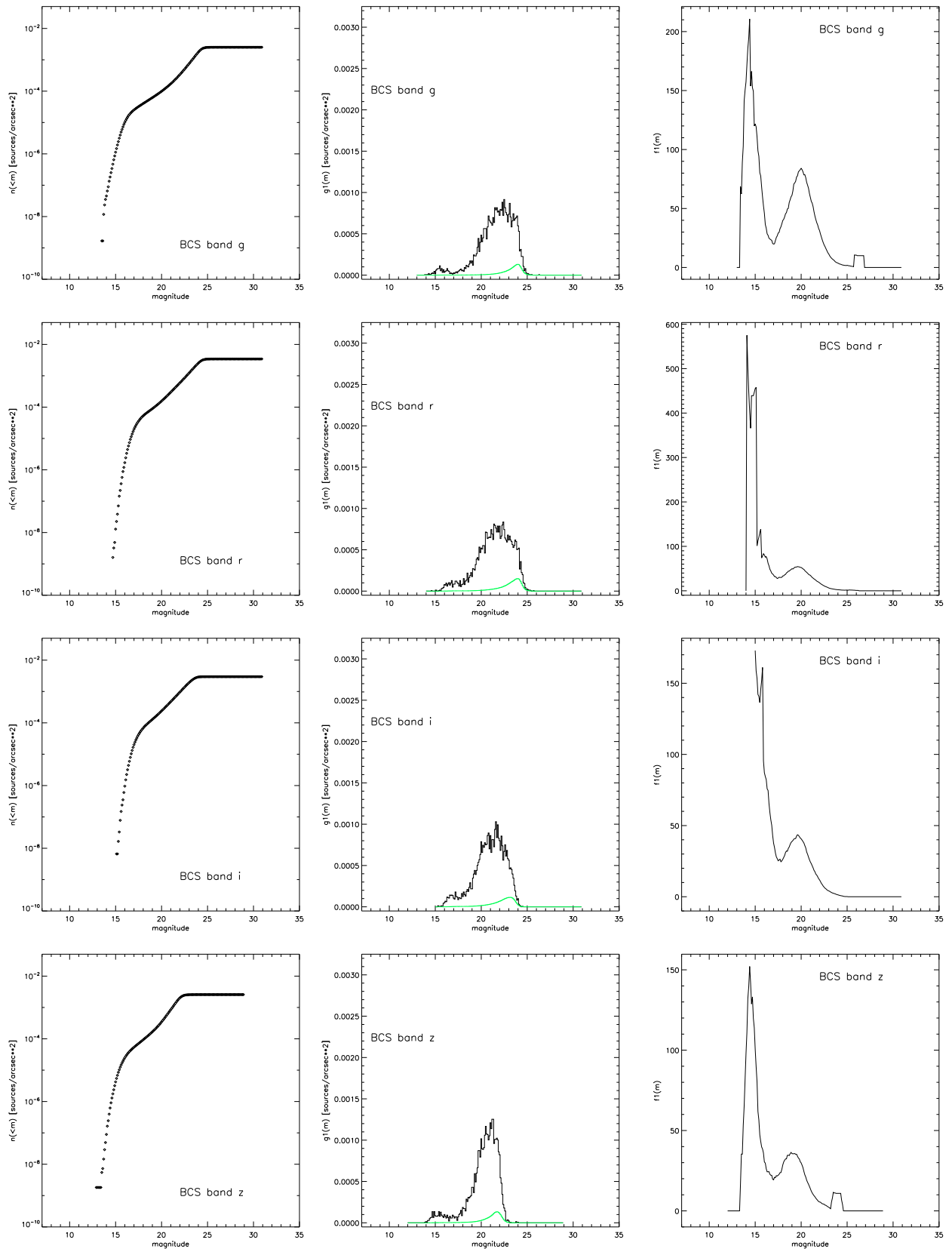


Fig. 23. Continuation of Fig. 16 for BCS bands

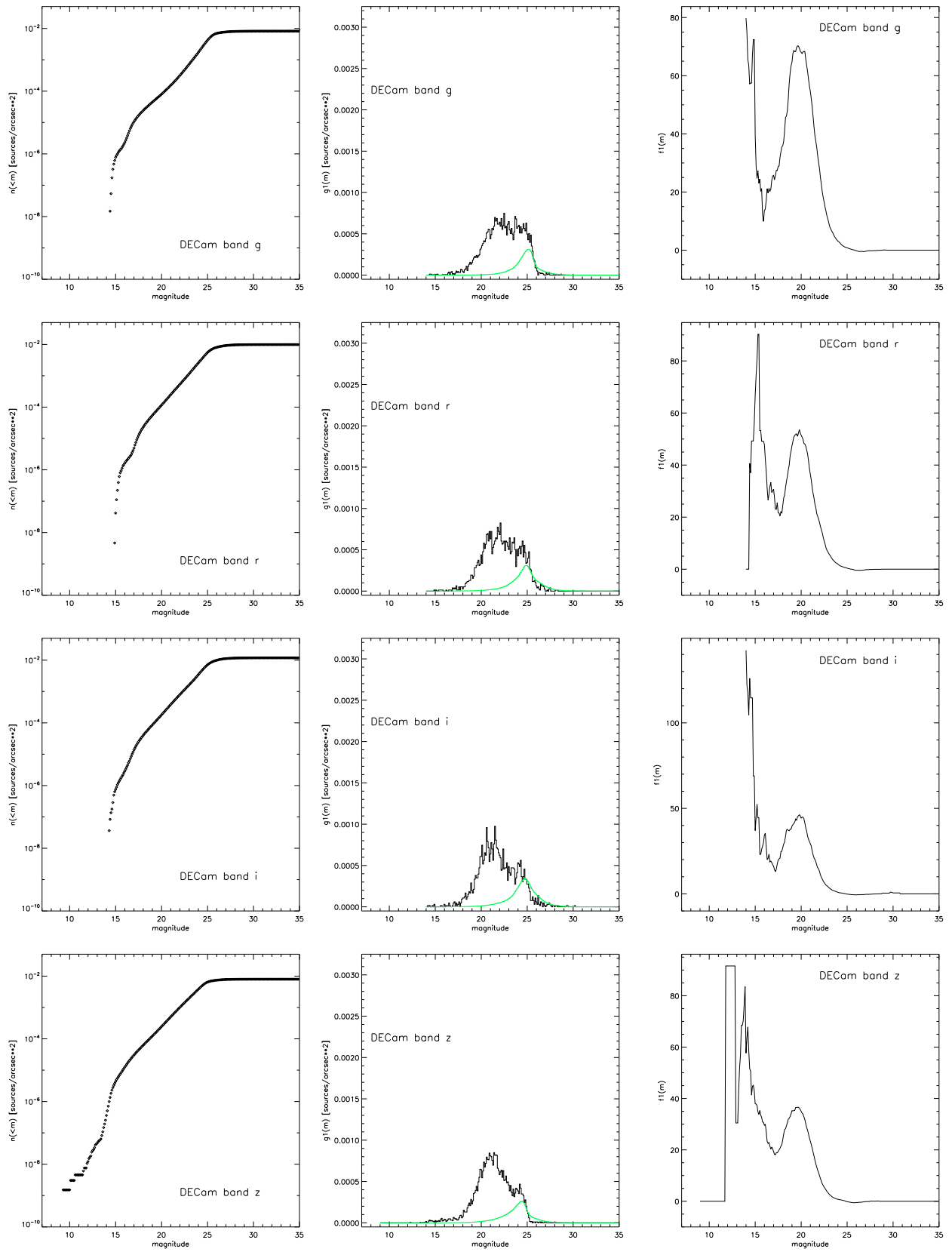


Fig. 24. Continuation of Fig. 16 for DECAM bands

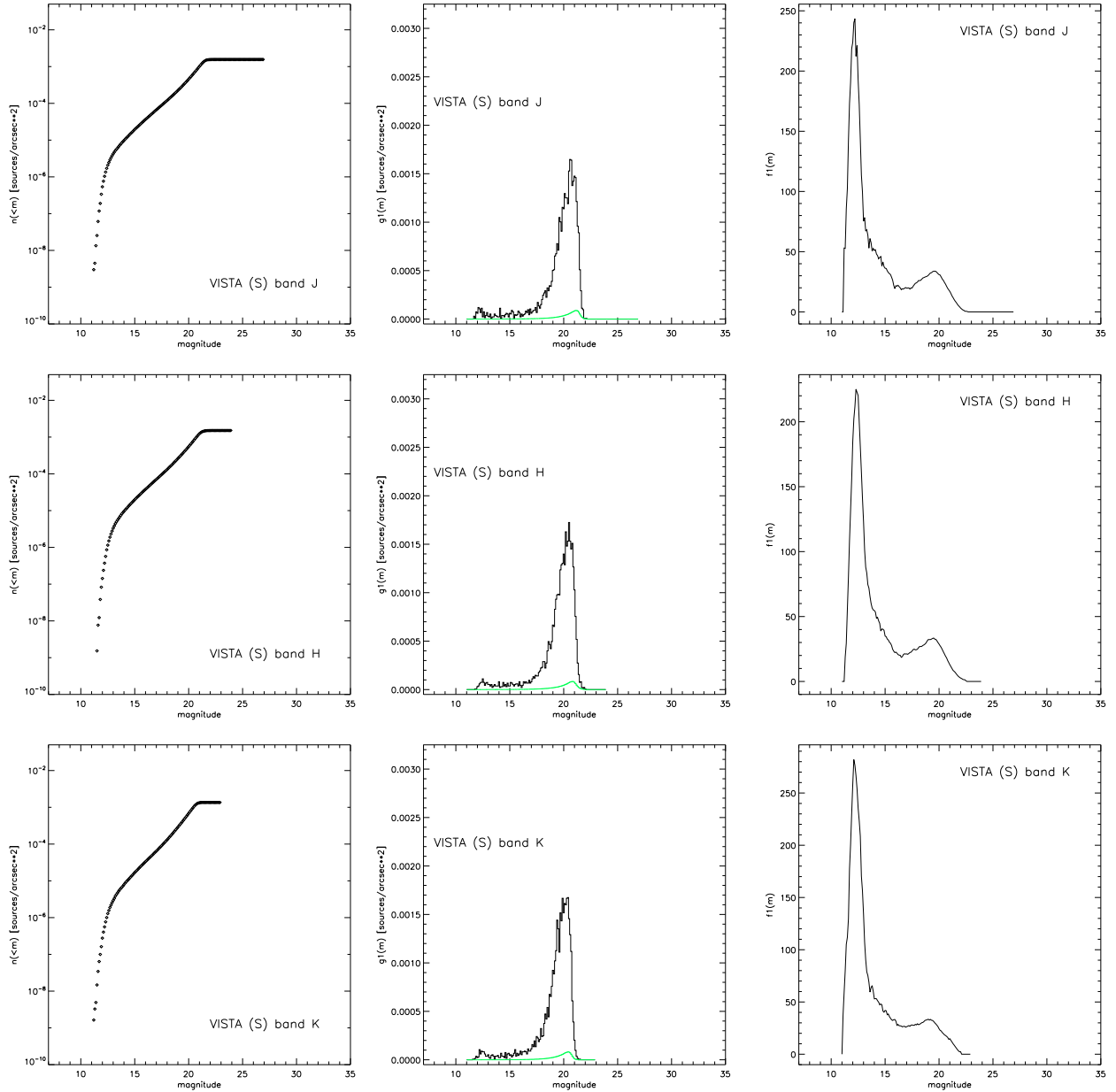


Fig. 25. Continuation of Fig. 16 for south VISTA bands

The ball diagrams in Fig. 28 report the number of occurrences of the various combinations of indicators.

Item	Probability	Likelihood ratio
Rank	1=average 2=best	3=average 4=best
Score	5=average 6=best	7=average 8=best

In general what ranks best for one indicator ranks best also in the other ones. The best (i.e. best band) rank is worse than the average one. What scores good for probability usually scores good also for LR, but not always (the P-based score is more relaxed than the LR-based one. Good scores are usually 1st rank, but, particularly for the

average ranks, several 1st rank counterparts score fair or bad.

The ranks and scores for the categories (optical, NIR, IR), for the individual surveys or for the bands are not unlike those for all bands (except of course they could be undefined where there is no measurement in the band).

7.5. Tying all together

All the previously described scores and order ranks are too many for practical use. I decided therefore to concentrate on just two classifications, one based on the best probability and one on the best LR of all bands, and generate

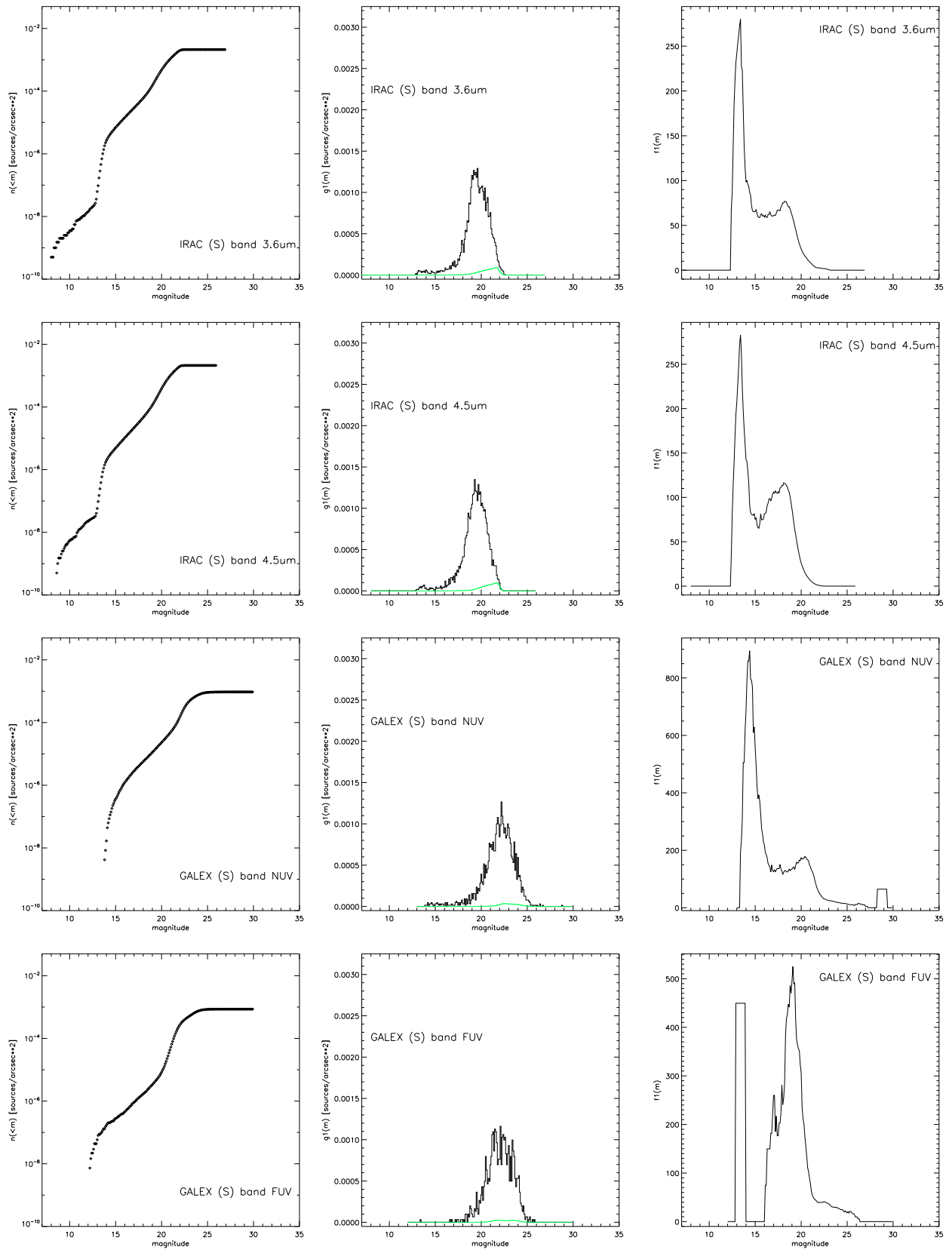


Fig. 26. Continuation of Fig. 16 for south IRAC and GALEX bands

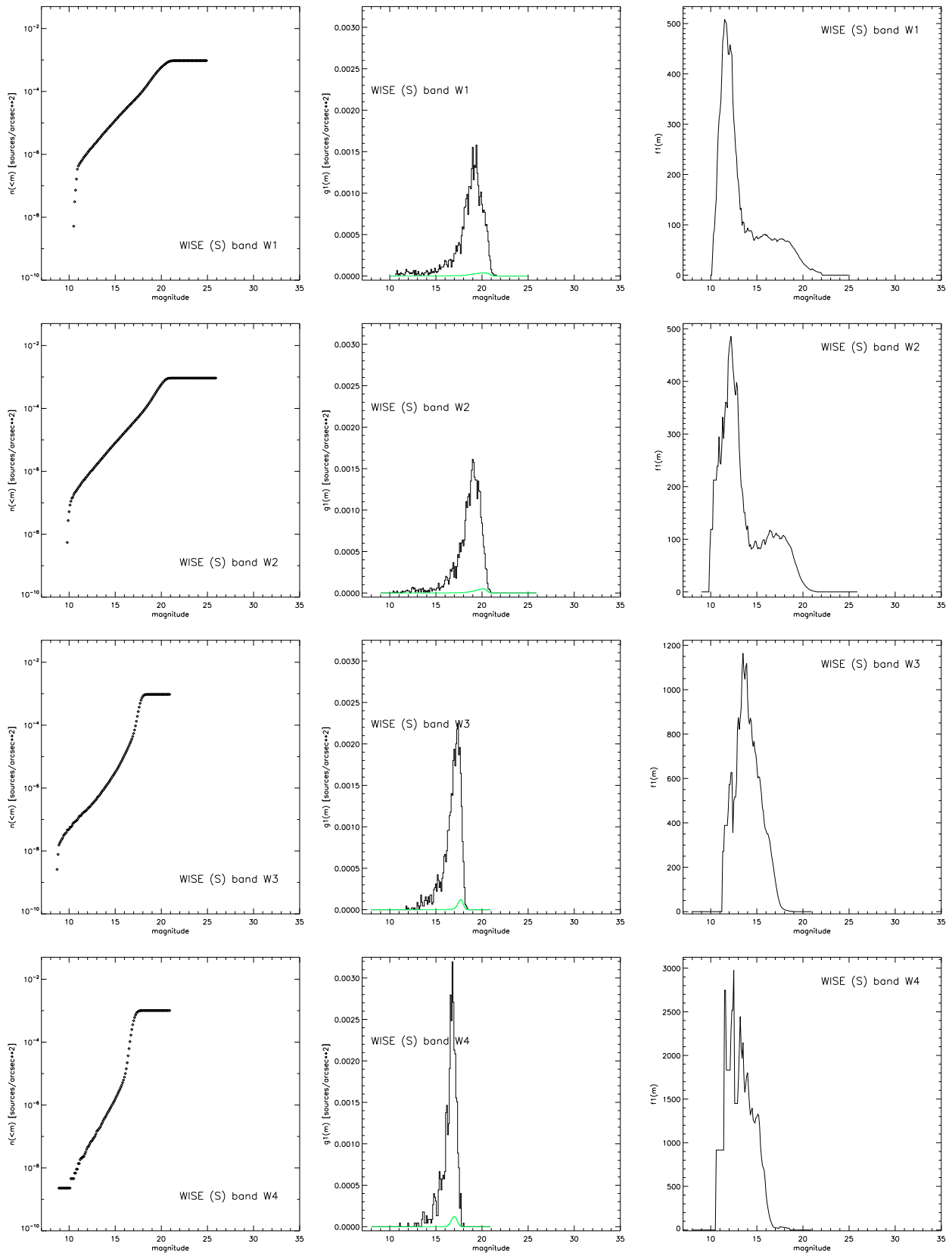


Fig. 27. Continuation of Fig. 16 for south WISE bands

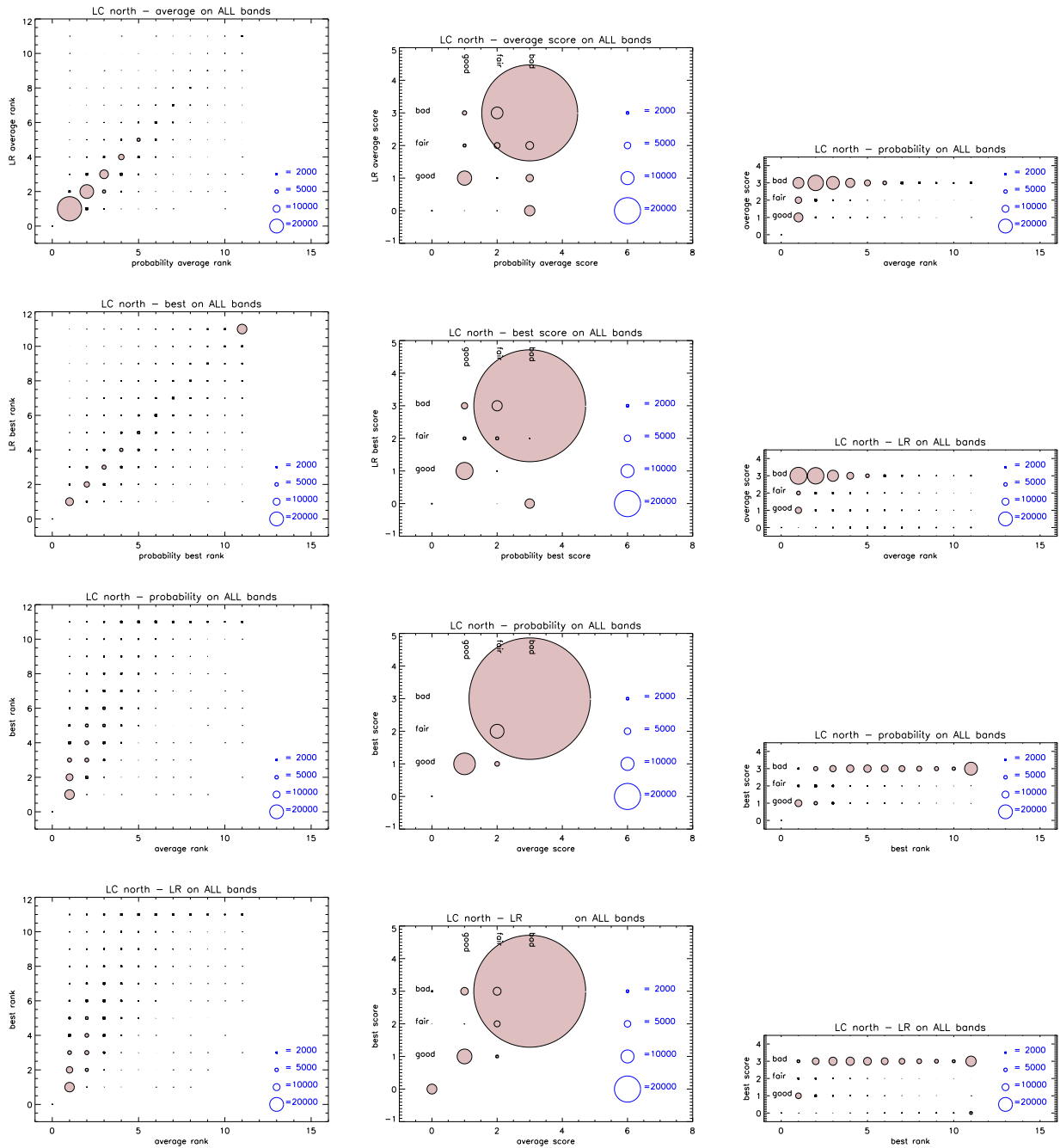


Fig. 28. Comparison of indicators for GCT *glor1cn*. Filled circles are proportional to the number of occurrences of the various combination (as indicated by the fiducial blue circles, scale may differ for some panels). The leftmost column panels report, from top to bottom: the LR average rank vs P average rank, the LR best rank vs P best rank, the best vs average rank for P, and the same for LR. The central column panels do the same for the scores (1=good, 2=fair, 3=bad, 0=unassigned). The rightmost column panels report, from top to bottom: average score vs rank for P, the same for LR, best score vs best rank for P and the same for LR

in parallel two "super-ranks", which in the remainder I'll call *P-rank* and *LR-rank*. To each of them I associate a *subrank* to allow to discriminate some details.

The definition is similar but not identical to the one used for XMM-LSS in Chiappetti et al. (2013), and is summarized here.

Rank	Subrank	Explanation
0		single counterparts
0	0	physically solitary
0	1	logically solitary
0	2	recovered reject
0	4	no counterpart (blank field)
1		multiple counterparts
1	0	best of ambiguous
1	1	marginally best ("pylo")
2		secondaries
2	1	second marginal ("pylo")
2	2	any other secondary
-1		rejected
-1	0	undefined P or LR
-1	-1	rejected as bad score

In practice I proceeded as follows, for each GCT, and separately based on probability P and on likelihood ratio LR.

- first of all a subrank=4 is assigned to the few cases where there is no catalogued counterpart in any band (*putative blank field* subject to visual verification). These are the same for P or LR, since they depend just on the absence of counterparts (P and LR are *undefined*). The blank field are considered within the firmly established counterpart case (rank=0).
- next a subrank=0 (within rank=0) is assigned to the relatively few cases where an X-ray source has a single counterpart set. These are termed *physically solitaires* and are the same for P or LR, since they depend just on the number of counterparts being equal to one. They are also irrespective of the score being good, fair or bad,
- next one rejects the relatively few cases where P or LR are unassigned for all bands (i.e. having a score zero). They are assigned rank=-1 and subrank=0. Cases like this occur when a source is e.g. detected in a single band, and no aperture magnitude is computed but only a mag_auto. They are usually photometric measurements with a `sflags=appended`.
- contextually one initially *rejects* (with rank=-1 subrank=-1) all the cases which are scored bad *both on the best and average* P (or respectively LR)
- at this stage it is possible that an X-ray source remains with only one counterpart which has not been rejected. This is considered a firmly established case (rank=0) assigned subrank=1, and termed *logically solitary*. By construction they score good or fair in P or respectively LR.
- At a later stage (after the ambiguity analysis described below) a recovery of rejects will be performed (see at end).
- One has to consider the case where an X-ray source has *more than one* counterpart sets, scored good or

fair, and perform an *ambiguity analysis*. This goes in several steps.

- One counterpart set per X-ray source will be considered *best though ambiguous*, and assigned rank=1.
- All other counterpart sets will be considered *secondaries* and assigned rank=2.
- In practice first one takes the object with the first order rank (i.e. lowest P or highest LR) for each X-ray source and provisionally assigns a subrank=9 (within rank=1).
- Then one takes the object with the second order rank (on P or LR) for each X-ray source (there will be at least one by construction) and provisionally assigns a subrank=10 (within rank=2).
- All what remains are *bona fide secondaries*, and assigned rank=2, subrank=2;
- Then for each X-ray source one compares the subrank=9 with its subrank=10. The former is considered *definitely better* if the ratio of P or LR with the latter (which is the immediately next choice) is greater than 10 (in the correct sense for P or LR).
- If it is, the subrank=9 is reflagged subrank=0 (*best of ambiguous*, i.e. it is highly likely to be the real counterpart), and the subrank=10 is reflagged subrank=2 (i.e. like *any other secondary*).
- If it is not, the subrank=9 and 10 are reflagged subrank=1, because the former is only marginally better than the latter. These are referred as *pylocatabatic*² cases, or "pylo" for short. There is always one pylo rank=2 associated to a pylo rank=1, while all the secondaries of a subrank=0 will have indifferently the same subrank=2.
- At the end the mentioned *recovery of the rejects* is performed. If an X-ray source is not in a blank field, and has no counterparts yet, the one with the best P or LR from the reject list is recovered, and assigned subrank=2. These cases could conservatively be considered unidentified sources, with a counterpart associated by chance (but could also have been rejected marginally, e.g. with a P=0.031 just above the threshold of 0.03).

Of course each X-ray source has a single preferred counterpart set which has rank between 0 and 1, plus possibly any number of rejects, plus (if ranked 1) one or more secondaries. All with variable subranks which might be used to take or leave each case.

Provisionally all P, LR, order ranks and scores are kept in an auxiliary side table offline while the two sets of rank and subrank have been moved to the main GCT, together

² The word "pylocatabasis" was created by Umberto Eco (Foucault's Pendulum, Binah 12, Sub umbra alarum tuarum) to designate *the art of being saved by a hair* (in Italian *scamparla per un pelo* or in better English *by the skin of one's teeth*, Job 19,20 ... perhaps Eco's translator should have used "odontoderma" ☺)

LCN		LR-based →	Total	rank=0					rank=1			identified	rank=2			reject	undef.	other
		P-based ↓		blank field	p.solitary	l.solitary	recovered	best amb.	best pylo		sec. pylo		other sec.					
rank	sub	Total	124112	9943	18	124	5905	3996	4225	842	3383	14168	8106	3383	4723	101838	313	101525
0		rank=0	6178	4843					29				1			1305		
0	4	blank field	18	18														
0	0	phys. solitary	124		124													
0	1	log. solitary	4197			2958	731											
0	2	recovered rej.	1839			35	971											
1		rank=1	7990	2811					3605				493			1081		
1	0	best ambig.	1927							493	219							
1	1	best pylo	6063							311	2582							
0/1		all identified	14168									11288						
2		rank=2	18993	713					587				7525			9568		
2	1	second. pylo	6063											2432	380			
2	2	other secondary	12330											521	4192			
-1		reject	91551	1576					4				87			89884		
-1	0	undef	313														313	
-1	-1	other	91238															89571

Table 4. Statistics of the ranking for GCT glor1cn

SFN		LR-based →	Total	rank=0					rank=1			identified	rank=2			reject	undef.	other
		P-based ↓		blank field	p.solitary	l.solitary	recovered	best amb.	best pylo		sec. pylo		other sec.					
rank	sub	Total	70854	12324	135	849	6930	4410	1844	315	1529	14168	2210	1529	681	54476	52	54424
0		rank=0	10026	8714					19				2			1291		
0	4	blank field	135	135														
0	0	phys. solitary	849		849													
0	1	log. solitary	6865			5193	1162											
0	2	recovered rej.	2177			31	1344											
1		rank=1	4142	1777					1549				246			570		
1	0	best ambig.	1055							197	107							
1	1	best pylo	3087							105	1140							
0/1		all identified	14168									12059						
2		rank=2	5728	429					273				1931			3095		
2	1	second. pylo	3087											1092	132			
2	2	other secondary	2641											198	509			
-1		reject	50958	1404					3				31			49520		
-1	0	undef	52														52	
-1	-1	other	50906															49468

Table 5. Statistics of the ranking for GCT glorsfn

with the new super-ranks described below. If there is a request, I could make the auxiliary table a member of the GCT, so one could access the additional columns by name (but there are some naming clashes which at the moment advised against this solution).

For each GCT one can provide a statistic of the ranking, which allows to spot the agreements and disagreements between the probability and likelihood methods, as shown in Tables 4 to 7.

Taking as example the first of such tables for the GCT glor1cn, the top-left cell on the diagonal lists the total number of candidate counterparts (124112), while the "to-

tal identified" value in the top column or in the leftmost column gives the number of X-ray sources (14168). By definition all are nominally identified, although 18 are blank fields, 47 of the 124 physical solitaires are scored bad both by P and by LR, and additional 22 are scored bad by LR, and more (3996) or less (1839) are flagged subrank=2, i.e. scored bad by LR or respectively by P, so could be considered dubious identifications.

The numbers along the diagonal are the cases which are classified exactly in the same way by P or by LR. The cell near the double rules indicates the counterparts which are preferred (rank 0 or 1) by both methods (P

LCS		LR-based →	Total	rank=0					rank=1			identified	rank=2			reject	undef.	other
		P-based ↓		blank field	p.solitary	l.solitary	recovered	best amb.	best pylo		sec. pylo		other sec.					
rank	sub	Total	8535	21	80	4937	3497	3353	794	2559	11888	5782	2559	3223	100997	2822	98175	
0		rank=0	4755	3713					4				-			1038		
0	4	blank field	21	21														
0	0	phys. solitary	80	80														
0	1	log. solitary	3418	2240 736														
0	2	recovered rej.	1236	6 630														
1		rank=1	7133	2809					2913				347			1064		
1	0	best ambig.	1839						462 171									
1	1	best pylo	5294						310 1970									
0/1		all identified	11888									9439						
2		rank=2	15927	747					436				5427			9317		
2	1	second. pylo	5294										1862 366					
2	2	other secondary	10633										410 2789					
-1		reject	90852	1266					-				8			89578		
-1	0	undef	2822													2822		
-1	-1	other	88030													86756		

Table 6. Statistics of the ranking for GCT glorlcs

SFS		LR-based →	Total	rank=0					rank=1			identified	rank=2			reject	undef.	other
		P-based ↓		blank field	p.solitary	l.solitary	recovered	best amb.	best pylo		sec. pylo		other sec.					
rank	sub	Total	10594	166	957	5545	3926	1294	265	1029	11888	1470	1029	441	35786	13	35773	
0		rank=0	8172	7121					2				-			1049		
0	4	blank field	166	166														
0	0	phys. solitary	957	957														
0	1	log. solitary	5371	3833 1162														
0	2	recovered rej.	1678	3 1000														
1		rank=1	3716	1848					1124				151			593		
1	0	best ambig.	1116						169 69									
1	1	best pylo	2600						90 796									
0/1		all identified	11888									10095						
2		rank=2	4993	468					168				1317			3040		
2	1	second. pylo	2600										777 101					
2	2	other secondary	2393										113 326					
-1		reject	32263	1157					-				2			31104		
-1	0	undef	13													13		
-1	-1	other	32250													31091		

Table 7. Statistics of the ranking for GCT glorsfs

or LR): 11288 i.e. 80%. The other cells provide a breakdown of more or less important discrepancies. For instance 2811 are ranked 1 (ambiguous) by P but are unambiguous (rank=0) by LR, while only 29 are ambiguous by LR and rank=0 by P: this is an unimportant discrepancy, the same counterpart set is chosen anyhow ... only the number of fair score secondaries is higher according to the less restrictive criteria based on P with respect to LR. Subrank variations are also unimportant. More critical and worth an inspection are the cases where one criterion takes a counterpart set as best (rank 0 or 1), and the other cri-

terion considers it a secondary or even a reject and hence prefers another counterpart set.

The situation for GCT glorsfn is cleaner (see Table 5). There are less candidate counterpart sets, and (hence) less secondaries and less ambiguities. There are more blank fields and solitaires. 85% of the best counterparts are the same for both P and LR methods.

For the southern GCT glorlcs (see Table 6) the statistics is similar to its equivalent northern case. 79% of the best counterparts are the same for both P and LR methods. There is a limited number of blank fields and

solitaries. There is a higher number (2822) of counterpart sets with both P and LR undefined.

The last GCT `glorsfs` (see Table 7) is also cleaner. 85% of the best counterparts are the same for both P and LR methods. The number of counterpart sets with both P and LR undefined is extremely limited.

7.5.1. Probability vs LR

In order to compare the performance of the probability-based ranking with the LR-based ranking, I have introduced in each GCT a *superrank* for each counterpart set, so defined:

- **0 (Firm primary)**
both P and LR ranks are between 0 and 1, i.e. the best counterpart set is the same for both criteria
- **1 (Possible primary)**
best counterpart according to one criterion, but secondary or reject for the other
- **2 (Firm secondary)**
secondary according to both criteria
- **3 (Possible secondary)**
secondary according to one criterion, but rejected for the other
- **-1 (Reject)**
anyhow rejected

In the northern area, GCT `glorlcn` has 11288 counterpart sets (unique for each X-ray source) which have *superrank*=0, i.e. both criteria give the same best counterpart in 80% of the cases. For 2880 X-ray sources (20%) the criteria differ and we have 5760 *superrank*=1 counterparts. 3923 X-ray sources have firm secondaries (total 7525), and 5497 have other secondaries (in total 7556 X-ray sources have some form of secondary).

GCT `glorsfn` has 12059 (85%) X-ray sources with the best counterpart common to both criteria, and 2109 (15%) X-ray sources with 4218 *superrank*=1 counterparts. A total of 3717 X-ray sources have some form of secondary (1638 with 1931 firm secondaries, and 2458 with others), i.e. this GCT is somewhat cleaner.

In the southern area GCT `glorlcs` has 9439 *superrank*=0 cases i.e. both criteria give the same best counterpart also in 80% of the cases. For 2449 X-ray sources (20%) the criteria differ and we have 4898 *superrank*=1 counterparts. 3155 X-ray sources have firm secondaries (total 5427), and 5264 have other secondaries (in total 6725 X-ray sources have some form of secondary).

GCT `glorsfs` has 10095 (also 85%) X-ray sources with the best counterpart common to both criteria, and 1793 (15%) X-ray sources with 3586 *superrank*=1 counterparts. A total of 3290 X-ray sources have some form of secondary (1171 with 1317 firm secondaries, and 2413 with others), i.e. this GCT is also somewhat cleaner.

One should trust the large majority of *superrank*=0 cases, and concentrate on the inspection

of *superrank*=1, and perhaps on the cases with a *rank*=1 and possible "pylo" secondaries.

7.5.2. SF vs LC comparison

The other comparison is the one between the LC-style and SF-style GCTs for the same area, exploiting the "unids" defined in 6.1.3. It is relatively easy to spot the cases where the unid is exactly the same, i.e. the counterpart sets are fully in common, identical and include the same objects in all the surveys.

This occurs for 60162 cases in the north (48% of the candidates in `glorlcn` and 84% of those in `glorsfn`), and for 41336 cases in the south (35% of the candidates in `glorlcs` and 84% of those in `glorsfs`).

It is less easy to study the cases where there is a partial match. I proceeded comparing the unid survey by survey and creating 8-character flags in the north (6-character in the south), for all cases which have a counterpart in a survey in common between the LC and SF GCT. The examples refer to the 6-character flags for simplicity. Excepting the cases with fully identical unids (flagged 'UUUUUU'), I obtain combinations like e.g. 'SS-zSz', which is interpreted as follows: S means the counterpart is the *same* for BCS, DECam and GALEX (and is not null); z means the counterpart is null in both GCTs for IRAC and WISE; - means the (VISTA) counterpart is different, which may mean both defined and different, or one null and the other defined.

The flags were generated separately in the direction LC to SF and SF to LC. There are several hundreds different flag combos, but they can be grouped according to various criteria.

A flag like 'zzzzzz' (all nulls) occurs only for the SF vs LC comparison, and correspond to the few cases (62 in the north and 95 in the south) of blank fields using SF counterparts (while they have a proper counterpart according to LC). Note that blank fields in both (unid 0:0:0:0:0:0) are considered a perfect match 'UUUUUU'.

A flag like 'SSSSSS' (all same) occurs in a few tens (north) to a couple hundreds cases (south) ... and one may ask why aren't they 'UUUUUU'? The answer is that ...they are same to some other but not all surveys are same to the same! *Consider for instance glorlcs 200015:1798:1680:31253:1620:765:908. It matches in 5 surveys glorsfs 200015:1798:1680:1685:1620:765:908, also flagged 'SSSSSS', but VISTA is different. VISTA 31253 matches in SF this 200015:0:0:31253:46824:0:15160, flagged 'zzSSzS', and the IRAC and WISE of the latter match two other 'zzzSzS' and 'zzz-zS' in LC.*

A flag like '-----' (all minus) occurs only for the LC vs SF comparison, in tens of thousands, and corresponds to counterpart sets present only in the LC GCTs.

One has then cases which contain only minus and z, only minus and S, only S and z, or all three cases. They

comparison →	glorlen									sfpointer count	glorsfn									lcpointer count	
	0 perf. match	1 close match	2 part. match	3 blank field	4 mix ambig.	5 ambig cpt.	6 other ambig.	8 uninteresting	9 LC only		0 perf. match	1 close match	2 part. match	3 blank field	4 mix ambig.	5 ambig cpt.	6 other ambig.	8 uninteresting	9 LC only		
superrank ↓	0	1	2	3	4	5	6	8	9	defined	74475	8283	2124	1434	62	82	19	-55	defined	64342	
0 firm primary	8283	31	437	-	961	10	18	712	836	defined	74475	8283	2124	1434	62	82	19	-55	defined	64342	
1 pos. primary	2253	897	211	-	522	-	5	825	1047	-99999	3375	2253	1187	625	-	150	1	2	-	-99999	6395
2 f. secondary	746	568	831	-	3094	-	15	984	1287	zero	46262	746	15	1145	-	25	-	-	-	zero	117
3 p. secondary	2213	519	498	-	1813	1	4	2155	2452			2213	3	763	-	145	1	1	-		
-1 reject	43327	1325	1811	-	7453	-	4	17860	18104			43327	11	2996	-	3182	-	4	-		

Table 8. Final statistics for the northern GCTs (in terms of counterparts, *not* X-ray sources).

comparison →	glorlcs									sfpointer count	glorsfs									lcpointer count	
	0 perf. match	1 close match	2 part. match	3 blank field	4 mix ambig.	5 ambig cpt.	6 other ambig.	8 uninteresting	9 LC only		0 perf. match	1 close match	2 part. match	3 blank field	4 mix ambig.	5 ambig cpt.	6 other ambig.	8 uninteresting	9 LC only		
superrank ↓	0	1	2	3	4	5	6	8	9	defined	51576	6529	2135	1072	95	58	136	20	50	defined	44601
0 firm primary	6529	16	289	-	556	83	202	697	1067	defined	51576	6529	2135	1072	95	58	136	20	50	defined	44601
1 pos. primary	1619	780	133	-	353	18	52	780	1163	-99999	2163	1619	1280	491	-	139	29	28	-	-99999	4398
2 f. secondary	567	489	491	-	1795	47	119	723	1196	zero	64928	567	4	719	-	6	20	1	-	zero	145
3 p. secondary	2199	579	465	-	1480	25	101	1749	2727			2199	4	672	-	134	15	18	-		
-1 reject	26995	1563	1019	-	5121	4	50	23952	30874			26995	4	1970	-	2101	4	30	-		

Table 9. Final statistics for the southern GCTs (in terms of counterparts, *not* X-ray sources).

can be grouped as follows introducing a *comparison* flag in each GCT, defined as follows (the numbering is intentionally not continuous):

- **0 (Perfect match)**
the unid is the same, and also same superrank, i.e. same counterpart and same classification
- **1 (Close match)**
the unid is the same, but superrank differs, i.e. same counterpart, different classification
- **2 (Partial match)**
Sz cases, unid partially matches (where not null)
- **3 (SF blank field)**
zzzzzz, blank field for SF, has a non null counterpart for LC
- **4 (Mixed ambiguous)**
Sz- cases, the mixed bag of all combinations
- **5 (Ambiguous in different counterparts)**
SSSSS, all counterparts not null, but may point to different candidates in each survey
- **6 (Other ambiguous)**
S- cases, unid matches in some surveys not in other
- **8 (Uninteresting)**
z- cases, unid is different or null
- **9 (LC only)**
-----, all surveys differ, items present only in LC GCTs

An example of Sz case is SF 200200:20101:18754:15302:16157:5807:0 flagged SSSSSz which is fully included in LC 200200:20101:18754:15302:16157:5807:10585 flagged SSSSS-. Another one is LC 200375:4492:4222:32022:3926:9919:0 flagged SSSSSz which matches SF 200375:4492:4222:0:3926:0:0 flagged SSzSz as well as 200375:0:0:32022:0:9919:0 flagged z-SzSz, which therefore could be merged. This justifies the idea that comparison=2 cases are the primary candidates for inspection for inclusion or merging.

An example of Sz- is SF 200121:5748:5402:5180:4964:0:2458 flagged -SSSz- which includes LC 200121:0:5402:5180:4964:0:0 flagged -SSSz- (remember that z means both are null, but minus allows one to be null). But LC 225387:40864:1697:1701:1636:771:0 flagged S----z matches SF 225387:40864:0:20963:25779:7251:0 flagged SzSSSs which matches back to several different LC entries. This justifies the idea that comparison=4 cases are more ambiguous.

Since for every area (north or south) we have two GCTs (LC and SF), one would like to associate fully or partially matching counterpart sets. I have defined a column *sfpointer* or *lcpointer* which allows to associate counterpart sets in either (LC and SF) GCT via their *seq*. The pointer can assume values

- 0 if the counterpart set is all different (present in one GCT only)
- positive and equal to the seq in the other GCT if the counterpart set matches fully or partially (in some surveys) the one in the other GCT
- -99999 if the counterpart set matches (partially) more than one entry in the other GCT, i.e. there are ambiguities

One can now combine the various flag and pointers to derive some final statistics, which are presented in Tables 8 and 9. The tables count the counterpart sets, but for the objects in the first line (superrank=0) this coincides with the number of distinct X-ray sources. The items coloured in green indicate the case where there is perfect agreement between all cases and methods, which can be trusted. **The items coloured in red are instead those definitely worth a visual inspection, with priority to those with comparison not zero.**

Concerning the pointers, we note that all comparison=5 cases are intrinsically ambiguous (the pointer is -99999), while all comparison=3 and comparison=8 have null pointers (i.e. are present in one GCT only), and correspond to (exhaust) the blank fields in the SF GCTs which aren't blank in LC GCTs (the other are included in the perfect matches).

58% of the northern X-ray sources and 55% of the southern are **uncontroversial** (superrank=0, comparison=0), to which one could add a further 10% of cases with superrank=1 but still comparison=0 (cases in red in the first column), while the remaining cases are worth inspecting.

7.5.3. SF vs SF comparison

I can in principle compare my "SF" GCTs with the *provisional* identification tables prepared by SF herself and referred in 3.1.2. Given that they are based on a Likelihood Ratio method, although the details are undocumented or at least unknown to me, and on the original photometric associations by SF, the obvious comparison is within the LR-based ranking of `glorsf?` vs `SF?ident`. Remember that, besides the details of the LR computation, there is another difference: the GCTs list *candidate counterparts*, possible primaries, secondaries and rejected, i.e. more than one for an X-ray source, while the identification tables have chosen *one counterpart* per X-ray source.

A more thorough comparison could be done when the identification tables are finalised. So far we limit to some basic statistics.

For reasons unknown to me, not all objects in the identification tables point to an `id` in the photometric tables. 784 X-ray sources in the north, and 1371 in the south have no `id`, although some of those have non-null α, δ coordinates in one of the two coordinate sets (the other one is set to -1). These objects may have a `flag` of -1 (also -

99 in the south), but sometimes a valid `flag` and LR. I presume, although this is not documented, that a `flag=1` indicates a good identification, and a `flag=3` a dubious one.

I compare the `flag` with my LR-rank and subrank

In the north 13384 sources (94% of the total, all those with a non-null `id` in the identification table) have a correspondence in `glorsfn` with the same `id` for the same X-ray source. 8981 (63%) are best for both (SF's `flag=1` and my LR-rank 0 or 1 irrespective of subrank); of the remaining `flag=1`, 483 (3%) are ambiguous secondary for me, 104 (less than 1%) are completely secondary for me, and 280 (2%) are rejected by me, i.e. for all those my primary is another object. 2812 (19%) have SF's `flag=3` and are OK for me (LR-rank 0 or 1), however of them 2353 (17%) are recovered rejects, subrank=2, so we sort of agree. Of the remaining `flag=3`, 710 (5%) are rejected by me, and just 12 are secondary for me (again for those my primary is another object).

Of the 784 (6%) cases with no `id` for SF, for 208 cases (2%) she assigns a `flag=1` to something else not in the photometric catalogues (?), and of these 5 for me are blank fields. For 336 (2%) she assigns a `flag=3` to something else (and 34 of these are blank fields for me). The rest has a negative `flag` (and include 96 of my blank fields).

In the south 10516 sources (88% of the total) have a correspondence in `glorsfs` with the same `id` for the same X-ray source. A single case for X-ray source 216651 is using an `id` I totally ignore. 7841 (70%) are best for both (SF's `flag=1` and my LR-rank 0 or 1 irrespective of subrank); of the remaining `flag=1`, 435 (4%) are secondary for me, and 322 (3%) are rejected by me, i.e. for all those my primary is another object. 1413 (12%) have SF's `flag=3` and are OK for me (LR-rank 0 or 1), however of them 1181 (10%) are recovered rejects, subrank=2, so we sort of agree. Of the remaining `flag=3`, 482 (4%) are rejected by me, and just 3 are secondary for me (again for those my primary is another object).

Of the 1371 (12%) cases with no `id` for SF, for 851 cases (7%) she assigns a `flag=1` to something else not in the photometric catalogues (?), and of these 44 for me are blank fields. For 367 (3%) she assigns a `flag=3` to something else (and 56 of these are blank fields for me). The rest has a negative `flag` (-1 or -99, and include 66 of my blank fields).

If we consider an agreement the cases flagged best by both, and the cases with `flag=3` and subrank 2, we can summarize that in 80% we agree, and we have to inspect the remaining 20%.

8. Visualization tool

In order to overlay the counterpart positions in the new multiwave catalogues onto the thumbnail images described in Report XVI (Chiappetti, 2014b), I created a

new version of the Java tool described in section 2 of Report XVI.

To download the new version (conventionally labelled as "version 32") just click on this URL in this PDF file: <http://sax.iasf-milano.inaf.it/~lucio/temp/Java/Alone/alone32.jar>. Refer to section 2.2 of Report XVI for testing. The invocation of the updated .jar file occurs via the following calls:

```
java -jar alone32.jar & for the northern area
java -jar alone32.jar S & for the southern area
```

Any argument other than a capital S defaults to the northern area.

For the operation and command reference of the tool refer to Report XVI. Note only the following differences:

- the catalogue menu in the control panel lists *only* the two north *or* the two south *NEW* catalogues (access to the old catalogue occurs via the old version 27 tool). Note that one can freely run more than one instance of the tool, in the same or different versions, e.g. to visualize variants side by side.
- the region table lists the counterparts in the GCT member tables (**north33or south33plus SF***) followed by a number of *external tables*
- the colour coding of the counterpart "regions" is different from the one in Table 5 of Report XVI and is explained below. The corresponding configuration file (**newtable.config**) is different from the one in 4.1.1 of Report XVI used by version 27.
- also the CGI files with the member table configuration described in 4.1.3 of Report XVI. The X11 window title bar may refer to "v32" or "v33" for the northern and southern version of the CGI.
- whenever the *size of the region* is conventional (function of magnitude and not an actual size on the sky), the best effort is made to always show a meaningful size (brightest magnitude in any band), although the "extra parameter" in the region table may refer to a fixed band (e.g. if a source is detected in *ugz* and brightest in *g*, the size of the region will be function of the *g* magnitude, but the "extra parameter" will report an undefined *i* = 99 magnitude).
- There is a new keyboard commands "**a**" which toggles the reduction of all circle and box regions by 1/3. This is useful to see better the underlying image, and to locate the counterpart "centre" closest to their position. However note that when toggle is in effect, those regions which do not have a conventional size (e.g. function of magnitude) but an actual size in arcsec are displayed with the *wrong size*.
- There is a new keyboard commands "**b**" which toggles the display of rejected counterparts (i.e. rank -1) as well as external objects (conventionally rank -9). The toggle has no immediate effect until regions aren't reloaded (so the usual sequence would be to issue **b** and **r** in sequence).

T	Table	colour	thickness	shape	size
X	X-ray merged	red	4	circle	maxdist or 4''
X	X-ray dup	red	1	circle	maxdist or 4''
X	X-ray soft pnt	pink	2	circle	radec_e
X	X-ray hard pnt	azure	2	circle	radec_e
X	X-ray soft ext	pink	2	circle	corerad
X	X-ray hard ext	azure	2	circle	corerad
North					
M	SFNgalax	violet	3	circle	f(mag)
e	galaxgr6	violet ⁻	1	circle	h(flux)
e	omsuss	violet ⁻⁻	1	circle	k(flux)
M	SFNcfht	cyan	3	circle	f(mag)
e	w1t7	cyan ⁻	1	circle	f(i)
M	SFNsdss	cyan ¹	3	circle	f(mag)
e	sdsssd10	cyan ¹⁻	1	circle	f(i)
M	SFNvista	yellow	3	circle	f(mag)
M	SFNwircam	yellow ¹	3	circle	f(K _s)
e	wircam	yellow ¹⁻	1	circle	f(K _s)
M	SFNukidss	yellow ²	3	circle	f(mag)
e	ukidssdr10	yellow ²⁻	1	circle	f(K)
M	SFNirac	orange	3	circle	f(mag)
e	irac2v0	orange ⁻	1	circle	f(mag ₃₆)
e	swiredr6	orange ⁻⁻	1	circle	g(flux)
e	SFNwise	orange ¹	3	circle	f(mag)
X	wise	orange ¹⁻	1	circle	f(W1)
S	gama	green	1	point	3 pix
S	vipers	bluish	2	circle	f(sel mag)
R	baran15	greenish	1	ellipse or box	
South					
M	SFSgalax	violet	3	circle	f(mag)
e	galaxgr6	violet ⁻	1	circle	h(flux)
M	SFSbcs	cyan ¹	3	circle	f(mag)
e	bcsru	cyan ¹⁻	1	circle	f(i)
e	bcs1mu	cyan ¹⁻⁻	1	circle	f(i)
M	SFSdecam	cyan	3	circle	f(mag)
e	decam	cyan ⁻	1	circle	f(i)
M	SFSvista	yellow	3	circle	f(mag)
M	SFSirac	orange	3	circle	f(mag)
e	ssdf2v9	orange ⁻	1	circle	f(mag ₃₆)
e	SFSwise	orange ¹	3	circle	f(mag)
X	wise	orange ¹⁻	1	circle	f(W1)
S	lidman15	green	2	point	3 pix
R	smolcic15	greenish	1	ellipse or box	
Auxiliary					
M	SF(N/S)zphot	orange-red	9	cross	0.3''
ME	ned	orange	1	point	4 pix
ME	simbad	orange/a	2	point	3 pix
ME	usno	white/blue	1	circle	f(R)
S	marseillespec	blue	1	point	2 pix
E	lyon	white	1	point	7 pix
E	agn1000(n/s)	whitegreen	4	box	f(i)
E	SF(N/S)ident	whitegreen	4	X cross	1''

Table 10. Region characteristics per table.

The first column (T) indicates the type of table (X=X-ray member; M=other member; non-members are E=external generic; S=spectroscopy; R=radio; e=old photometric tables, indented). A single indent is for the direct correspondent of the newer SF tables. Double indent for other tables in same band. Auxiliary tables are present for both the north and south area, unless they appear after an horizontal rule in the north or south sections. Thickness goes from 4 (thick) to 1 (thin). Superscripts indicates different shades of the quoted colour, and a superscript minus a fainter shade.

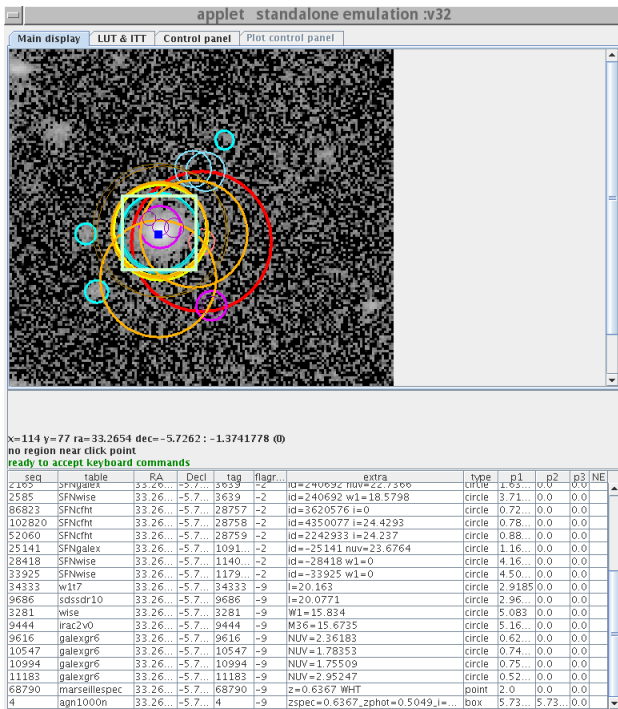


Fig. 29. Screenshot of the tool for one of the northern 1000 brightest AGN

The colour coding of the regions is described in Table 10. It has been updated in such a way to use the following criteria:

- X-ray region codes are unchanged vs Report XVI
- other codes are in a sort of wavelength order, UV is violet, optical is cyan, NIR is yellow-green, IR is orange. Radio is green. Stuff to be somehow emphasized is white.
- the SF member tables use a **thick brighter** colour. The equivalent older tables use a thinner and fainter shade of the same colour (note that when a region is selected/highlighted however it becomes thicker)
- X-ray and radio table use regions of "physical" size; spectroscopy and external tables usually use just a square marker; photometric tables use a circle function of the magnitude $f(\text{mag})$.
- For VIPERS (spectroscopy) table a conventional circle corresponding to the selected target magnitude is used. Objects from the 1000 brightest AGN catalogues i.e. Paper VI (Fotopoulou et al., 2016) use a square box of conventional size based on the i mag of the counterpart in the paper.
- For radio catalogues there may be two regions: a rectangular box with the nominal position errors in all cases, and an ellipse of size and position angle from the catalogue for resolved objects.
- The SF?zphot members (objects with photometric redshift) are currently represented with a thick small cross of fixed size, orange-red. Since virtually all (SF)

counterparts have a photometric redshift, such cross marks the "centre" of the counterpart set elements.

- The *provisional SF?ident* preferred counterparts, *one per source*, are currently marked by an X cross of fixed size.

Colours reported in Table 10 do not consider those used for highlighting as described in Report XVI.

The scaling $f(\text{mag}) = (26 - \text{mag})/2$, with a lower limit of $0.1''$, occurs on the *brightest* magnitude in any band which is *not undefined*. (In practice due to the way URL-encoded strings are passed from CGI to servlet to mysql this requires some tricks out of the scope of this document).

As said above, the tool displays primarily the member tables (X-ray or photometric), and also a choice of external tables. SIMBAD, NED and USNO have a *double role of member and external*: the objects associated as members as described in 6.1.4 appear twice, as part of a counterpart set and as external; the others appear only as external. Lyon XXLDB, spectroscopy and radio tables and currently SF?ident are external tables provided for convenience. Further external table are a *choice* of older photometric tables (those originally released within the DB and coming from public or authorized private sources), and are supplied to allow comparison. The preliminary homogenized tables supplied by SF in 2014 (Chiappetti, 2014a) and never released within the DB remain accessible *only* via the old version 27 tool (Chiappetti, 2014b).

9. Quick user's guide

First of all note that the NEW* views list *all candidate counterparts* within $10''$, not just the best candidates. They are working tools to validate the results and inspect the most critical (or ambiguous) cases. In a future clean counterpart tables, with just selected candidates, like the ones referred in 3.1.2 might be released.

Proceed as follows in the Milan database:

- scan the first columns of the "Tables" page until the category titled 'XXL merged tables (views)'
- tick one of the tables NEWSFN, NEWSFN, NEWLCN or NEWLCS. (clicking on the table name will instead open an help page with the names of the virtual columns)
- if one is interested in accessing parameters like magnitudes or photometric redshifts, one must scroll the page down to the very bottom and tick the box labelled 'Show member tables (and data products!) also'
- One shall then use the 'Advanced' query tab
- In the 'Column list' menu on the left one shall then see first the view virtual columns and next the long list of all member table columns.
- Let us assume one wants to select for an X-ray source all the aperture magnitudes and the photometric redshift: one could select (control-clicking on them) columns Xseq, SFNsds.s.magu to SFNsds.s.magz,

- then all `SFNcfht.mag*` and so on for the other surveys until `SFNwise.magw4`. Then add the redshift `SFNzphot.zphot`.
- Either control-click all columns you wish and then click on 'Insert columns' to get them in the order of appearance in the menu. Or click and insert one at a time for a custom order.
- If e.g. you then want to add the eventual identifier in SIMBAD, add column `simbad.id`
 - If you want the probability and LR based ranks, add columns `Prank` and `LRrank`
 - If you want to see also the common photometric id, you should manually make sure the column is named `glorsfn.id` or equivalent (for the view `NEWXXY` the GCT is called `glorxxy` in lower case (XX = SF or LC, Y = N or S))
 - then in the 'Selection condition' write something like `Xseq=200002` and submit the query
 - This way you can get all details for all counterparts of a given X-ray source, like the one for which you saw a summary in the visual tool described in 8.
 - If you want to avoid seeing the rejected counterparts, add to the 'Selection condition' a clause like `and (Prank between 0 and 2 or Lrank between 0 and 2)`. Mind the parentheses on the boolean expression !
 - Of course you are not obliged to query on `Xseq`, you could query for all non-rejected counterparts of any source using `Prank between 0 and 2 or Lrank between 0 and 2`
 - Or if you decide to follow the Probability way or the LR way, you can take a single, preferred counterpart per X-ray source with a clause like e.g. `Prank between 0 and 1` (or conversely `LRrank between 0 and 1`)
 - If you want to know whether the P and LR based rankings agree or not, you can add to the list the column `superrank` (see pag. 40 in 7.5.1) or use it in a condition
 - If you want to restrict the case to specific counterparts, you can add additional clauses.
 - E.g. `SIMBADseq is not null` will return only counterparts found in SIMBAD
 - Or `SFvista is not null` will return only counterpart set with a VISTA counterpart.
 - A clause `superrank=1` will return all cases in which the best counterpart selected by probability and the best counterpart selected by LR differ (one should inspect them and decide which way is the best)
 - A clause `Prank=0 and Psubrank=2` (or equivalent for LR rank) will single out the worse candidates (those which were recovered rejects, they would have been rejected by score, were not the case that this way the X-ray source will not have any counterpart)
 - A clause `Prank=0 and Psubrank=4` (or equivalent for LR rank) will locate the blank fields (X-ray sources without apparent counterpart)
 - A clause `Prank=1` (or equivalent for LR rank) will locate the cases with possible secondary counterparts
 - Similarly one can use clauses on column `comparison` (see pag. 41 in 7.5.2). to locate cases where the SF and LC views differ.
 - In this case one can get the `seq` of the corresponding entry in the other table: for instance from `NEWSFN`, hidden column `glorsfn.lcpointer` is the corresponding `NEWLCN.seq`. This might come useful for instance if one keeps side by side two instances of the visual tool described in 8
 - Do not forget that after executing a query, if you do 'View data' and then, in the resulting window, go to the rightmost column 'Object related files', you can download the thumbnail images used by the tool in section 8 or get direct access the NED or SIMBAD page if any.
- This is just a taste of the possible combinations. After all, *there is no royal way* and you have to read carefully the present report and experiment yourself.
- Acknowledgements. I acknowledge conversations with M.Polletta about the Likelihood Ratio method. I acknowledge the *huge* work done by S.Fotopoulou to prepare the datasets described in this work and released to XXL in February 2016, as well as her careful reading of the drafts of this report, and the supply of various inputs. I remind that the full catalogues cover much more than the immediate surroundings of the XXL X-ray source. Its usage beyond the direct XXL needs shall be negotiated with S.Fotopoulou in order to ensure due credit is given to her and no data rights are violated. The provisional reference to be cited is Fotopoulou et al. (2016).
- ## References
- Bongiorno, A. & Brusa, M., 2014, GAMA-XXL match notes, 15/11/2014 (internal note)
- Brusa, M., Zamorani, G., Comastri, A., et al. 2007, ApJS, 172, 353
- Chiappetti, L., 2013, The XXLN and XXLS catalogues, Preliminary release for internal use, XMM-LSS Internal Report N. 12-Mi (Report XII)
- Chiappetti, L., Clerc, N., Pacaud, F., et al. 2013, MNRAS, 429, 1652
- Chiappetti, L., 2014, 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)
- Downes, A. J. B., Peacock, J. A., Savage, A., & Carrie, D. R. 1986, MNRAS, 218, 31
- Fotopoulou, S., Pacaud, F., Paltani, S., et al. 2016, A&A, 592, A5 (arXiv:1603.03240)
- Sutherland, W., & Saunders, W. 1992, MNRAS, 259, 413