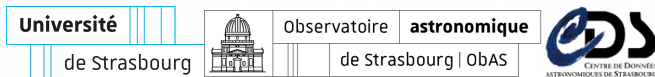


Outils de cross-correlation : accès aux catalogues et cross-identifications

F.-X. Pineau¹ et l'équipe du CDS

¹ CDS, Observatoire Astronomique de Strasbourg

Transient Sky 2020 National Workshop - 2nd Edition
Montpellier, 4-6th June, 2018



□ Plan

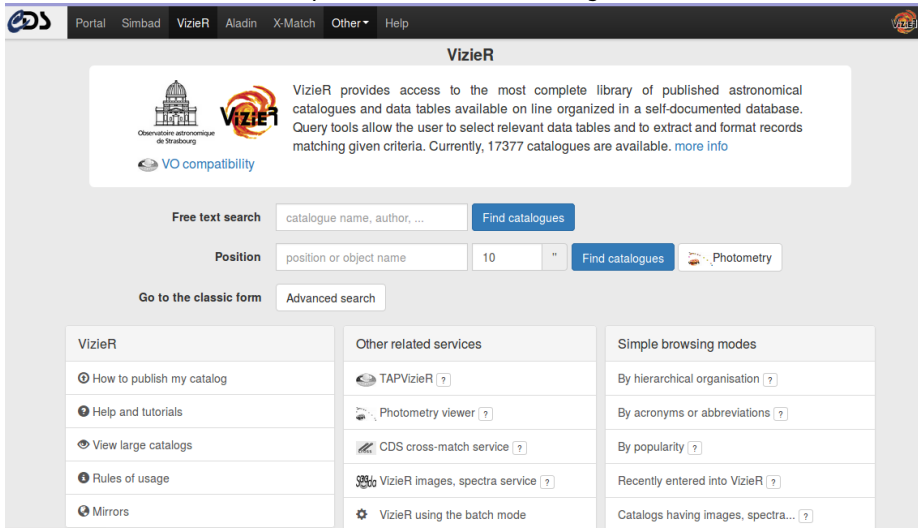
- Accès aux catalogues du CDS
 - ▶ tables normales, grandes tables
- Cross-match simple
 - ▶ service de xmatch du CDS
- Cross-match complexe
 - ▶ outil ARCHES
- Prise en compte de paramètres photométriques
 - ▶ estimation de vraisemblances photométriques par noyau
- Développements exploratoires:
 - ▶ Index spatio-temporel basé sur HEALPix
 - ▶ Test de Apache Spark




Accès aux catalogues



□ VizieR: interface web


<http://vizier.u-strasbg.fr/>




CDS Portal Simbad **VizieR** Aladin X-Match Other Help 

VizieR






  VizieR provides access to the most complete library of published astronomical catalogues and data tables available on line organized in a self-documented database. Query tools allow the user to select relevant data tables and to extract and format records matching given criteria. Currently, 17377 catalogues are available. [more info](#)

 [VO compatibility](#)

Free text search

Position  [Photometry](#)

Go to the classic form

| VizieR | Other related services | Simple browsing modes |
|---|--|--|
| How to publish my catalog |  TAPVizieR ? | By hierarchical organisation ? |
| Help and tutorials |  Photometry viewer ? | By acronyms or abbreviations ? |
| View large catalogs |  CDS cross-match service ? | By popularity ? |
| Rules of usage |  VizieR images, spectra service ? | Recently entered into VizieR ? |
| Mirrors |  VizieR using the batch mode | Catalogs having images, spectra... ? |

VizieR: interface web

Portal Simbad VizieR Aladin X-Match Other Help

VizieR

[Simple Target](#) [List Of Targets](#) [Fast Xmatch with large catalogs or Simbad](#)

Target Name (resolved by [Sesame](#)) or Position: Target dimension:

Radius Box size

1./345/gaia2 Gaia DR2 (Gaia Collaboration, 2018) [acknowledge](#) [Similar Catalogs](#) [2018A&A.in.pre...](#) [ReadMe+ftp](#)
[and cite Gaia DR2](#) [TimeSerie](#)

1./345/gaia2 Gaia data release 2 (Gaia DR2). (Download all Gaia Sources as VOTable, FITS or CSV [here](#). Query from the command line using [find_gaia_dr2](#) available in [cdsclient](#) [here](#))
(original column names in green) (1692919135 rows)

[Submit](#) [Reset All](#)

[Simple Constraint](#) [List Of Constraints](#)

Query by [Constraints](#) applied on Columns (Output Order: + -)

Standard Original

| Show | Sort | Column | Clear | Constraint | Explain (UCD) |
|-------------------------------------|--------------------------|-----------|----------------------|------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> | DR2Name | <input type="text"/> | (char) | Unique source designation (unique across all Data Releases) (Gaia DR2 NNNNNNNNNNNNNNNNNNNNN) (designaion) (Note J1) (meta.id) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | RA_ICRS | <input type="text"/> | deg | ⁽ⁱ⁾ Barycentric right ascension (ICRS) at Ep=2015.5 (ra) (pos.eq.ra:meta.main) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | e_RA_ICRS | <input type="text"/> | mas | Standard error of right ascension (e_RA*cosDE) (ra_error) (stat.error:pos.eq.ra) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | DE_ICRS | <input type="text"/> | deg | ⁽ⁱ⁾ Barycentric declination (ICRS) at Ep=2015.5 (dec) (pos.eq.dec:meta.main) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | e_DE_ICRS | <input type="text"/> | mas | Standard error of declination (dec_error) (stat.error:pos.eq.dec) |
| <input type="checkbox"/> | <input type="checkbox"/> | SolID | <input type="text"/> | | Solution Identifier (solution_id) (Note G1) (meta.id:meta.version) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Source | <input type="text"/> | | ⁽ⁱ⁾ Unique source identifier (unique within a particular Data Release) (source_id) (Note G2) (meta.id:meta.main) |
| <input type="checkbox"/> | <input type="checkbox"/> | RandomI | <input type="text"/> | | ⁽ⁱ⁾ Random index used to select subsets (random_index) (Note 2) (meta.code) |
| <input type="checkbox"/> | <input type="checkbox"/> | Epoch | <input type="text"/> | yr | [2015.5] Reference epoch (ref_epoch) (meta.ref.time.epoch) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Plx | <input type="text"/> | mas | ⁽ⁿ⁾⁽ⁱ⁾ Absolute stellar parallax (parallax) (pos.parallax) |

VizieR: interface web

CDS Portal | Portal | Simbad | **VizieR** | Aladin | X-Match | Other | Help

VizieR Send to VO tools

VO Table (bin-64)

XML + CSV (Aggregates)

Search Criteria

Search by table

Keywords (back)

Table compatible

HTML Table (Add)

HTML with Checkboxes

Detailed results

Table separated-Values

Table separated-Values

Table separated-Values

Table separated-Values

asci text/plain Query

asci table

Preferences

asci - with Checkboxes

max: 50

HTML Table

All columns

Compute

Submit

Mirrors

CDS, France

VizieR

Show the target form


Show constraint information




The 3 columns in **color** are computed by VizieR, and are **not part of the original data**.

/345/gaia2 [Gaia DR2 \(Gaia Collaboration, 2018\)](#) [2018A&A.in.prep...](#) [ReadMe+ftp](#)

[Post annotation](#) Gaia data release 2 (Gaia DR2). (**Download** all Gaia Sources as VOTable, FITS or CSV [here](#). **Query from the command line using find_gaia_dr2 available in cdsclient [here](#)**)

(original column names in green) (1692919135 rows)



 [start Aladin Lite](#)  [plot the output](#)  [query using TAP/SOL](#)

| Full | RA_ICRS deg | e_mas | DE_ICRS deg | e_mas | Source | Plx mas | e_mas | pmRA mas/yr | e_mas | pmDE mas/yr | e_mas | Dup |
|------|-----------------|--------|-----------------|--------|---------------------|------------|--------|----------------|-------|----------------|-------|-----|
| 1 | 300.29012518483 | 0.0400 | +30.95345885708 | 0.0575 | 2030849314289726080 | 0.3475 | 0.0622 | -4.772 | 0.092 | -5.037 | 0.124 | 0 |
| 2 | 300.28908599963 | 1.3677 | +30.95346547158 | 3.3967 | 2030849314254048512 | | | | | | | 0 |
| 3 | 300.29055144910 | 0.5894 | +30.95545880490 | 1.0960 | 2030849314254044544 | -1.3396 | 1.0439 | -1.766 | 1.568 | -6.348 | 3.161 | 0 |
| 4 | 300.29129691607 | 0.1716 | +30.95493850348 | 0.2531 | 2030849314247900800 | 0.5268 | 0.2652 | -3.742 | 0.397 | -3.538 | 0.543 | 0 |
| 5 | 300.29257452331 | 0.2802 | +30.95636129921 | 0.4217 | 2030849314289910400 | 0.6584 | 0.4342 | -3.307 | 0.668 | -5.233 | 0.916 | 0 |
| 6 | 300.29265439369 | 0.0992 | +30.95814047164 | 0.1426 | 2030849314247910528 | -0.0388 | 0.1563 | -1.804 | 0.234 | -5.760 | 0.312 | 0 |
| 7 | 300.29004853598 | 0.3818 | +30.95778899474 | 0.6715 | 2030849314253018624 | -0.4933 | 0.6914 | -3.508 | 1.061 | -4.629 | 1.845 | 0 |
| 8 | 300.29106674969 | 0.3056 | +30.95908906574 | 0.5313 | 2030849314289910656 | 0.0137 | 0.4913 | -3.829 | 0.727 | -5.378 | 1.140 | 0 |
| 9 | 300.29293541515 | 0.1896 | +30.95924558954 | 0.2715 | 2030849309955520256 | 0.3134 | 0.2998 | -2.301 | 0.436 | -5.497 | 0.591 | 0 |
| 10 | 300.28572825258 | 0.1209 | +30.95793369503 | 0.1762 | 2030849309955515520 | 0.5069 | 0.1826 | -2.751 | 0.278 | -5.713 | 0.358 | 0 |
| 11 | 300.28648445997 | 0.0589 | +30.95995021886 | 0.0914 | 2030849314289736960 | 0.3610 | 0.0929 | -2.399 | 0.134 | -4.502 | 0.179 | 0 |
| 12 | 300.28967329915 | 0.1432 | +30.96107288738 | 0.2087 | 2030849314289733376 | 0.0262 | 0.2228 | -3.419 | 0.321 | -5.093 | 0.427 | 0 |
| 13 | 300.28936831030 | 0.1935 | +30.96065909881 | 0.2880 | 2030849314247956864 | 0.5088 | 0.3294 | 3.659 | 0.454 | -2.240 | 0.558 | 0 |
| 14 | 300.29832585088 | 1.5340 | +30.95489004117 | 4.2115 | 2030849348613768960 | | | | | | | 0 |
| 15 | 300.29933699892 | 0.2293 | +30.95500530256 | 0.2830 | 2030849348607556480 | -0.0130 | 0.3488 | -2.068 | 0.485 | -4.813 | 0.617 | 0 |

□ VizieR: accès scriptable

<http://cds.u-strasbg.fr/resources/doku.php?id=cdsclient>

Python-cdsclient

Table of Contents

- Python-cdsclient
- CDSclients
- vizquery Package Installation
- Package Contents
- Proxys and firewall

The Python cdsclient package gather scripts to query large tables : wise, 2mass, sdss, Gaia, ... It needs no special installation (except that python version 2 or 3 must be installed).

1. Download :  Python cdsclient package
2. Install :

```
tar -xzf python-cdsclient.tar.gz
cd python-cdsclient
```

and test:

```
./find_2mass.py -h
```

the package includes:

- vizquery.py: a generic script which enables to query VizieR using the VizieR identifiers (ex 2mass=II/246)


```
List big catalogues : vizquery.py -l
Get columns information for 2mass (=II/246) : vizquery.py -source=II/246 -l
Query Gaia (I/337/gaia) around M1 (18arcsec) : vizquery.py -source=I/337/gaia -c=M1 -c.rs=10
Get hipparcos (HIP=1) in votable : vizquery.py -source=I/239/hip_main -mime=votable -out.max=50 "HIP=1"
```

- dedicated scripts like find_allwise.py, find_gaia_dir2.py, etc

```
Query 2mass arround M1: find_2mass.py M1
Query sdss12 arround '217.488910+36.086880': find_sdss-dr12.py "217.488910+36.086880"
```

You can also query with constraints (see `-help` to list the constraints available (case sensitive))

News

-  CDS last news

Developers

Java

- Multi-Order Coverage Map
- Unit Conversion
- Astronomical coordinates and proper motions
- VOTable Parsing
- VOSpace
- UWS Library
- ADQL Library
- TAP Library
- Aladin source code
- Downloads

New Interfaces and

Interactions

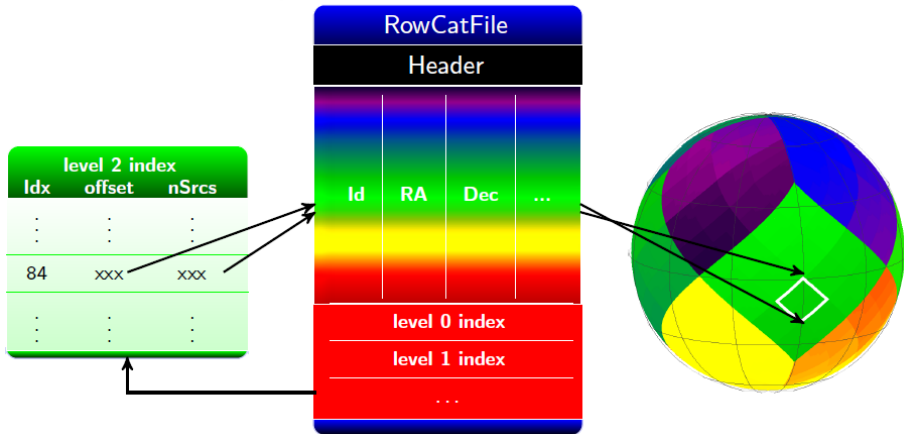
- Android & IOS devs
- HTML5, WebGL
- Multitouch screens, etc.

Unix/Linux

- CDSClient
- Anafile
- Downloads
- H3C PostgreSQL library

□ VizieR: accès scriptable

Particularité des grands catalogues: pas en base!



Format binaire spécifique

□ Vizier: accès scriptable

<http://tapvizier.u-strasbg.fr/adql/>

The screenshot shows the Tap Vizier web interface. At the top, there is a navigation bar with links for Portal, Simbad, Vizier, Aladin, X-Match, Other, and Help. The main heading is "Tap Vizier". A warning message states: "Warning: Crossmatch turning [note](#)". Below this, a blue box contains the text: "The TAPVizieR service provides [VizieR tables using ADQL](#) (a SQL extension in Astronomy)." To the right, a "Documentations" section lists links for "About TAP Vizier" and "ADQL documentation & examples".

The main area prompts the user to "Type your ADQL Query in the bottom area or try an example" with a dropdown menu showing "2mass_". Below this is a yellow box for "Search tables" with a "Go" button and instructions: "Search by catalog, author's name, word(s) from title, position (resolved by [Sesame](#)), ... e.g : Veron, 2Mass, redshift , M31...".

To the right, a "Favorite tables available to construct queries" section includes a note: "* You can not make query on more than two tables. * Selected tables are automatically stored locally." Below this is a "Construct your query" button and an "Upload your data" section with a "Name File/Url" field and a note: "* to use an uploaded table in the query, you must prefix its name with TAP_UPLOAD (i.e. TAP_UPLOAD.myTable)."

The bottom section contains a text area with the following ADQL query:

```
1 --2mass_around_M1
2 -- the objects from 2mass around M1 within 1 arcmin
3 SELECT "II/246/out".raj2000, "II/246/out".dej2000, "II/246/out".Jmag, "II/246/out".Kmag, "II/246/out".Jmag-"II/246/out".Kmag as j_k
4 FROM "II/246/out"
5 WHERE 1=CONTAINS(POINT('ICRS', raj2000, dej2000), CIRCLE('ICRS', 83.633083, 22.0145,1/60))
```


At the bottom, there are controls for "Query name" (set to "2mass_around_M1"), "Output format" (set to "csv"), and buttons for "Run", "Quickview", "Reset", and "Test".



Cross-match simple

XMatch service

http://cdsxmatch.u-strasbg.fr


Portal Simbad VizieR Aladin X-Match Other Help

CDS X-Match Service X-match Tables management Documentation Login Preferences Register


Choose tables to cross-match

VizieR SIMBAD My store VizieR SIMBAD My store

[The SDSS Photometric Catalog, Release 9 \(Adelman-McCarthy+, 2012\)](#)
794,013,950 rows





[2MASS All-Sky Catalog of Point Sources \(Cutri+ 2003\)](#)
470,992,970 rows



Visualize and manage your cross-match jobs

List of X-match jobs

| Table 1 | Table 2 | Options | Begin | Status | Actions |
|----------|---------|---|---------------------|---|---|
| SDSS DR9 | 2MASS | fixed radius  radius: 5 arcsec area: All sky | 03/11/2016 at 14:15 | completed  | <input type="button" value="Get result"/> |

Job executed in **10min11s**
3min40s to correlate
6min31s to generate file
Result: **66,006,865** rows (19.3 GB)

□ XMatch service

Example

Example

Using `curl` to match several FITS file with Simbad in Bash

```
for f in file1 file2 file3 file4; do \  
  curl -X POST -F request=xmatch \  
    -F cat1=@$f.fits -F colRA1=RAJ2000 -F colDec1=DEJ2000 \  
    -F cat2=simbad \  
    -F distMaxArcsec=25 \  
    -F RESPONSEFORMAT=csv \  
    http://cdsxmatch.u-strasbg.fr/xmatch/api/v1/sync \  
  > $f_vs_simbad_25arcsec.csv \  
done
```

Other languages

For Python, Ruby and Java, see here:

<http://cdsxmatch.u-strasbg.fr/xmatch/doc/xmatch-API-usage-examples.html>

XMatch service

<http://cdsxmlmatch.u-strasbg.fr/xmatch/api/v1/sync>

The screenshot shows the TOPCAT software interface. The title bar reads 'TOPCAT'. The menu bar includes 'File', 'Views', 'Graphics', 'Joins', 'Windows', 'VO', 'Interop', and 'Help'. The toolbar contains various icons for file operations, data visualization, and analysis. A red arrow points to the 'X' icon in the toolbar, which is used for XMatch queries. The main window is divided into two panes: 'Table List' on the left and 'Current Table Properties' on the right. The 'Table List' pane shows a single table entry: '1: l%239%hip_main.fits'. The 'Current Table Properties' pane displays the following information:

- Label: l%239%hip_main.fits
- Location: /home/pineau/data2/HADOOP/l%239%hip_main.fits
- Name: l/239/hip_main
- Rows: 118,218
- Columns: 14
- Sort Order: ↑
- Row Subset: All
- Activation Action: (no action) Broadcast Row

At the bottom of the interface, there is a 'SAMP' section with 'Messages:' and 'Clients:' fields. The 'Messages:' field is empty, and the 'Clients:' field shows two active client icons.

Usage statistics

- Web Interface (removing internal usages)

| year | #IPs | #Jobs | | #Links | | Outputs size | |
|-------------|------|-------|-----------|---------|--------------|--------------|-------------|
| | | | /day | Billion | M/day | TB | GB/day |
| 2017 | 1706 | 8873 | 33 | 47.9 | 179.4 | 13.61 | 52.2 |
| 2016 | 1923 | 11102 | 30 | 37.9 | 104.0 | 10.1 | 28.4 |
| 2015 | 1194 | 7406 | 20 | 20.3 | 55.7 | 5.0 | 14.0 |
| 2014 | 1136 | 5909 | 16 | 25.6 | 70.2 | 6.6 | 18.5 |
| 2013 | 1081 | 5407 | 14 | 5.0 | 13.7 | 1.2 | 3.4 |
| 2012 | 535 | 3699 | 10 | 11.5 | 31.4 | 2.7 | 7.5 |
| 2011 | 96 | 409 | 7 | 3.7 | 67.3 | 0.83 | 15.5 |

- XXX: computed on incomplete years

Usage statistics

- Synchronous HTTP API (removing internal usages)

| year | #IPs | #Jobs /day | #Links | | #Positions (TOPCAT) | |
|-------------|------|---------------|---------|-------------|---------------------|-------------|
| | | | Billion | M/day | Billion | M/day |
| 2017 | 1664 | 1250 | 3.4 | 12.2 | 4.1 | 14.8 |
| 2016 | 1765 | 889 | 2.5 | 6.7 | 4.3 | 11.9 |
| 2015 | 1099 | 580 | 2.4 | 6.6 | 3.0 | 8.3 |
| 2014 | 406 | 49 | 0.6 | 1.6 | 0.3 | |
| 2013 | 46 | | 0.1 | | | |

- XXX: computed on incomplete years



Cross-match complexe

The ARCHES tool

<http://serendib.unistra.fr/ARCHESWebService/index.html>

ARCHES X-MATCH TOOL Anonymous Web form



[Info about this page.](#)

Remote directory

Upload a file:

Parcourir...

Aucun fichier

File list

```
3xmme_uniquesources.  
2mass.174.10491_7.223  
sdss9.174.10491_7.223  
galex5ais.174.10491_7.
```

X-match script

Script examples

Xmatch galex/sdss/2mass in a cone, with proba

Type, modify or copy/paste here the xmatch script to be executed:

```
1 #####  
2 # Name: galex_sdss_2mass.xmls  
3 # Description: Perform a probabilistic xmatch between galex, sdss and 2mass  
4 # in a given cone of 12 arcminutes. Data is downloaded from VizierR.  
5 # Input files: none  
6 # Output files:  
7 # - galex.vot: galex data  
8 # - sdss9.vot: sdss data  
9 # - 2mass.vot: 2mass data  
10 # - galex_sdss_2mass.vot: cross-match result  
11 # WARNING: the result may not be symmetric using successive full joins  
12 #####  
13  
14 # Load galex data from VizierR  
15 get VizierLoader tablename=ll/312/ais mode=cone center="174.10491 +7.22343" radius=12.0arcmin allcolumns  
16 set pos ra=RAJ2000 dec=DEJ2000  
17 set poserr type=CIRCLE param1=0.6  
18 set cols objid,.*J2000/(e_)?[FN]UV/  
19 prefix galex_  
20 save galex.vot votable  
21  
22 # Load sdss data from VizierR  
23 get VizierLoader tablename=V/139/sdss9 mode=cone center="174.10491 +7.22343" radius=12.1arcmin allcolumns  
24 where mode==1 && e_RAJ2000>0.0 && e_DEJ2000>0.0 && mag<23  
25 set pos ra=RAJ2000 dec=DEJ2000  
26 set poserr type=RCD_DEC_ELLIPSE param1=e_RAJ2000 param2=e_DEJ2000  
27 set cols objid,.*J2000/(e_)?[FN]UV/
```

Submit

The ARCHES tool

1 sur 28

Zoom automatique

Introduction
Going beyond the two-catalogue case
Simplifying assumptions
▼ Notations and links with catalogues
Notations
Classical positional errors in catalogues
▼ Candidates selection: the match
Estimation of the real position given n observations
Candidates selection criterion
Iterative form: catalogue by catalogue
Iterative form: by groups of catalogues
Summary and Interpretation
Comment on the "Bayesian cross-match" of Budavari2008
▼ Hypotheses from combinatorial considerations
Generalities
▼ Possible combinations and the Bell number
Two-catalogues case: two hypotheses
Three-catalogues

A&A 597, A89 (2017)
DOI: 10.1051/0004-6361/201629219
© ESO 2017

Astronomy & Astrophysics

Probabilistic multi-catalogue positional cross-match

F.-X. Pineau¹, S. Derriere¹, C. Motch¹, F. J. Carrera², F. Genova¹, L. Michel¹, B. Mingo³, A. Mints^{4,5},
A. Nebot Gómez-Morán¹, S. R. Rosen³, and A. Ruiz Camuñas²

¹ Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, 67000 Strasbourg, France
e-mail: francois-xavier.pineau@astro.unistra.fr
² IFCA (CS-IC-UC), Avenida de los Castros, 39005 Santander, Spain
³ Department of Physics & Astronomy, University of Leicester, Leicester, LE1 7RH, UK
⁴ Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
⁵ Max-Planck Institute for Solar System Research, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany

Received 30 June 2016 / Accepted 23 August 2016

ABSTRACT

Context. Catalogue cross-correlation is essential to building large sets of multi-wavelength data, whether it be to study the properties of populations of astrophysical objects or to build reference catalogues (or timeseries) from survey observations. Nevertheless, resorting to automated processes with limited sets of information available on large numbers of sources detected at different epochs with various filters and instruments inevitably leads to spurious associations. We need both statistical criteria to select detections to be merged as unique sources, and statistical indicators helping in achieving compromises between completeness and reliability of selected associations.

Aims. We lay the foundations of a statistical framework for multi-catalogue cross-correlation and cross-identification based on explicit simplified catalogue models. A proper identification process should rely on both astrometric and photometric data. Under some conditions, the astrometric part and the photometric part can be processed separately and merged a posteriori to provide a single global probability of identification. The present paper addresses almost exclusively the astrometric part and specifies the proper probabilities to be merged with photometric likelihoods.

Methods. To select matching candidates in n catalogues, we used the Chi (or, indifferently, the Chi-square) test with $2(n-1)$ degrees of freedom. We thus call this cross-match a χ -match. In order to use Bayes' formula, we considered exhaustive sets of hypotheses based on combinatorial analysis. The volume of the χ -test domain of acceptance – a $2(n-1)$ -dimensional acceptance ellipsoid – is used to estimate the expected numbers of spurious associations. We derived priors for these numbers using a frequentist approach relying on

□ The ARCHES tool

| Algorithm | param | #tbl | prop.mot. | index struct. |
|--------------------|--------|------|-----------------|---------------|
| chi2 (χ^2) | proba | 2 | l^1, r^2, b^3 | M/TM-tree |
| proba2_vx | proba | 2 | no (?) | M-tree |
| proba3_vx | proba | 3 | no (?) | M-tree |
| proba4_vx | proba | 4 | no (?) | M-tree |
| probaN_vx | proba | n | no (?) | M-tree |
| knn | k+dist | 2 | r, b | kd/M/TM-tree |
| cone | dist | 2 | l, r, b | kd/M/TM-tree |
| mec ¹ | dist | n | no (?) | kd/M-tree |
| ext_l ¹ | r | 2 | no | M-tree |
| ext_r ² | r | 2 | no | M-tree |
| ext_b ³ | r | 2 | no | M-tree |
| ... | ... | ... | ... | ... |

XMatches are chainable: 1 χ^2 xmatch of 4 tables = 3 χ^2 xmatches of 2 tables!

4 to 11 joins ($LIR\bar{F}\bar{L}\bar{I}\bar{R}'I'R'F'$) are supported according to the algorithm.

- ¹ l: left table contains extended objects or proper motions;
- ² r: right table contains extended objects or proper motions;
- ³ b: both left and right tables contain extended objects or proper motions.

□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

- ★ AB (H_0)

- ★ A_B

A
•

•
B

- For 3 catalogues

- ▶ 5 hypothesis

A
•
• B • C

□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

- ★ AB (H_0)

- ★ A_B

A
•

•
B

- For 3 catalogues

- ▶ 5 hypothesis

A
•
• B • C

□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

- ★ AB (H_0)
- ★ A_B



- For 3 catalogues

- ▶ 5 hypothesis



□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

- ★ AB (H_0)
- ★ A_B



- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ A_BC
- ★ A_B_C
- ★ A_B_C



□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

- ★ AB (H_0)
- ★ A_B

A
•
B
•

- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ AC_B
- ★ A_BC
- ★ A_B_C

A
•
B • C

□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

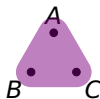
- ★ AB (H_0)
- ★ A_B



- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ AC_B
- ★ A_BC
- ★ A_B_C



□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

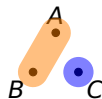
- ★ AB (H_0)
- ★ A_B

A
•
•
B

- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ AC_B
- ★ A_BC
- ★ A_B_C



□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

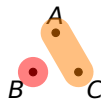
- ★ AB (H_0)
- ★ A_B

A
•
B
•

- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ AC_B
- ★ A_BC
- ★ A_B_C



□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

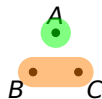
- ★ AB (H_0)
- ★ A_B



- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ AC_B
- ★ A_BC
- ★ A_B_C



□ Hypotheses

To compute Bayes probabilities, we MUST consider all possible hypothesis.

- Law of total probabilities:

$$\sum_{i=1}^n p(H_i) = 1$$

- For 2 catalogues

- ▶ 2 hypothesis

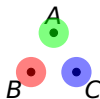
- ★ AB (H_0)
- ★ A_B

A
•
•
B

- For 3 catalogues

- ▶ 5 hypothesis

- ★ ABC (H_0)
- ★ AB_C
- ★ AC_B
- ★ A_BC
- ★ A_B_C



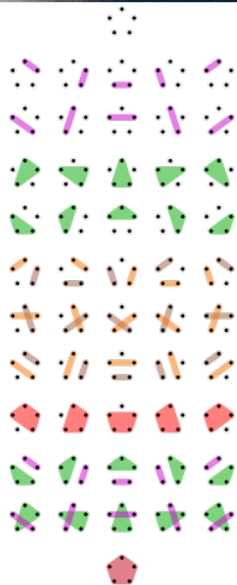
□ Hypotheses

- We generalised for n catalogues
- The number of hypothesis to be tested is given by the BELL number

Table : Values of the seven first BELL numbers

| n | 2 | 3 | 4 | 5 | 6 | 7 |
|-------|---|---|----|----|-----|-----|
| B_n | 2 | 5 | 15 | 52 | 203 | 877 |

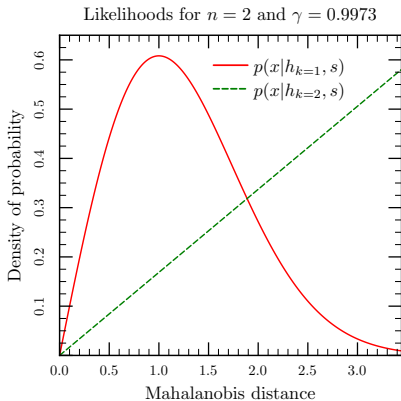
- ▶ n number of catalogues
- ▶ $n=5$ catalogues \rightsquigarrow 52 probabilities to be computed
- \Rightarrow Combinatorial explosion!



Bayesian probabilities

Summary for 2 catalogues

- 2 hypotheses
- 2 likelihoods
 - ▶ $H_0 = AB$: $p(x|H_{AB})$, Chi of 2 dof
 - ▶ $H_1 = A_B$: $p(x|H_{A_B})$, Poisson 2D
- 2 priors based on geometrical estimates



Bayesian probabilities

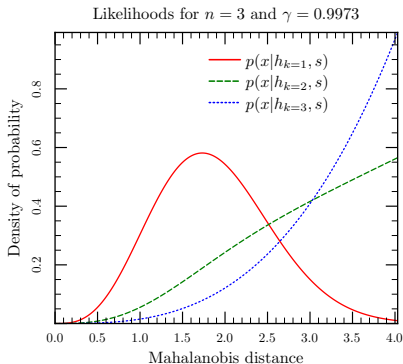
For 3 catalogues

- 5 hypotheses

- ▶ ABC , 1 real source
- ▶ AB_C , 2 real sources
- ▶ AC_B , 2 real sources
- ▶ A_BC , 2 real sources
- ▶ A_B_C , 3 real sources

- 3 likelihoods

- ▶ 1 likelihood by number of real source
- ▶ $p(x|H_{ABC}) = \chi_{dof=4}(x)/\gamma$: Chi 4 dof
- ▶ $p(x|H_{AB_C}) = p(x|H_{AC_B}) = p(x|H_{A_BC})$
- ▶ $p(x|H_{A_B_C}) = 4x^3/k_\gamma^4$: Poisson 4D





Prise en compte de paramètres photométriques

□ XMatch example

- Roughly reproducing Salvato et al. (2018) results (J/MNRAS/473/4937/xmmslew2) with the ARCHES tool + CDS classification (prototype) service
- Write and ARCHES cross-match script

```
# Load and set the XMM data to be cross-matched
```

```
get FileLoader file=XMMSL2_exgal_fewcol_2017APR12.fits
```

```
where RADEC_ERR < 10.0
```

```
set pos ra=RA dec=DEC
```

```
set poserr type=CIRCLE param1=RADEC_ERR/sqrt(2)
```

```
set cols *
```

```
prefix x
```

```
# Load and set the AllWISE data to be cross-match
```

```
get FileLoader file=candidate_ALLWISE_counterparts_unique_2017APR12.fits.gz
```

```
where eeMaj < 0.75
```

```
set pos ra=RA dec=DEC
```

```
set poserr type=ELLIPSE param1=eeMaj param2=eeMin param3=eePA
```

```
set cols *
```

```
prefix w
```

```
# Perform the cross-match, add the angular distance and save the result
```

```
xmatch probaN_v1 joins=I completeness=0.9973 area_w=0.01851769294883401575
```

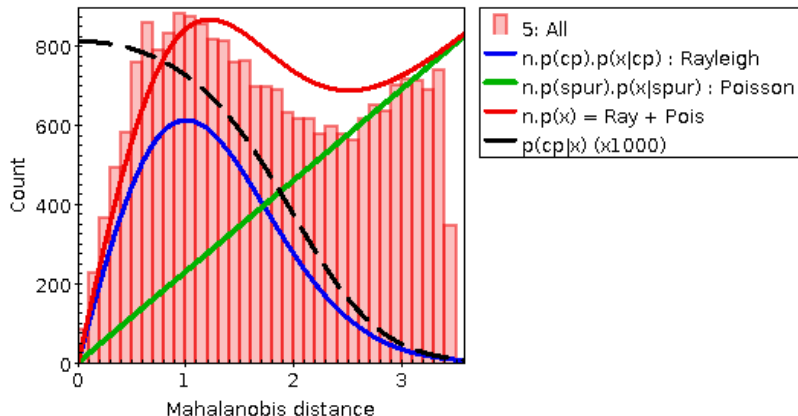
```
area_x=0.01851769294883401575 area_xw=0.01388
```

```
merge dist mec
```

```
save xmmslew2_vs_allwise.fits fits
```

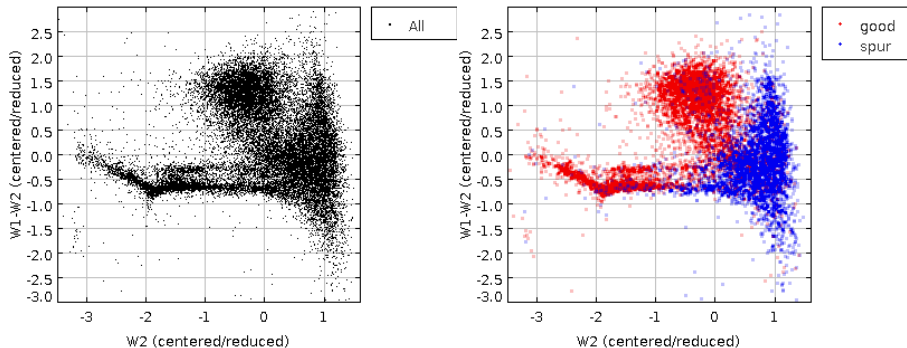
□ NWAY / ARCHES: result

- Cross-match result: 23 813 associations
- Number of spurious matches ($n.p(spur)$) overestimated (?)



□ NWAY / ARCHES: photometry

- From Salvato et al. (2018): W2 vs W1-W2
- Learning samples arbitrary defined:
 - ▶ Good: $d < 6''$ && $\text{RADEC_ERR} < 8$ && $\text{proba_xw} > 0.75$
 - ▶ Spurious: $d > 13''$ && $\text{proba_xw} < 0.05$



□ Classification service

- 3 CSV files: all matches, good matches, spurious matches
- Each file contains: *id, w2, w1 – w2*
- Using the CDS prototype service:

```
# Put the data files into the distant server
```

```
./classif.bash put good slew_vs_allwise.good.csv # 4565 rows
```

```
./classif.bash put spur slew_vs_allwise.spur.csv # 3082 rows
```

```
./classif.bash put data slew_vs_allwise.all.csv # 23799 rows
```

```
# Performs the classification of the data and save the result
```

```
./classif.bash kdc samplepoint -k 75 -p good:0.425\;spur:0.575 -ho > result.csv
```

```
# Ask for the confusion matrix by self-classifying the LS
```

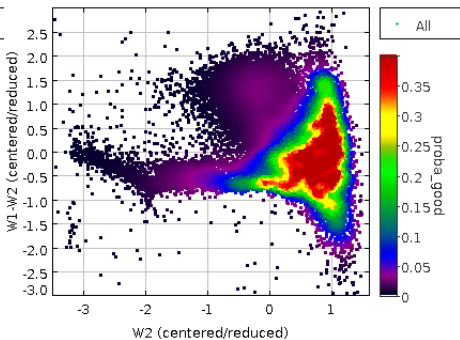
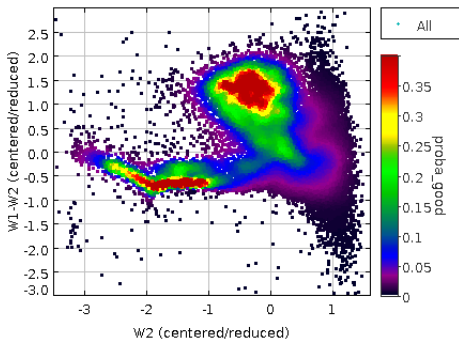
```
./classif.bash kdc samplepoint -k 75 -p good:0.425\;spur:0.575 -cr
```

- Confusion matrix:

| actual \ predicted | good | spurious |
|--------------------|--------|----------|
| good | 85.96% | 14.04% |
| spurious | 9.73% | 90.27% |

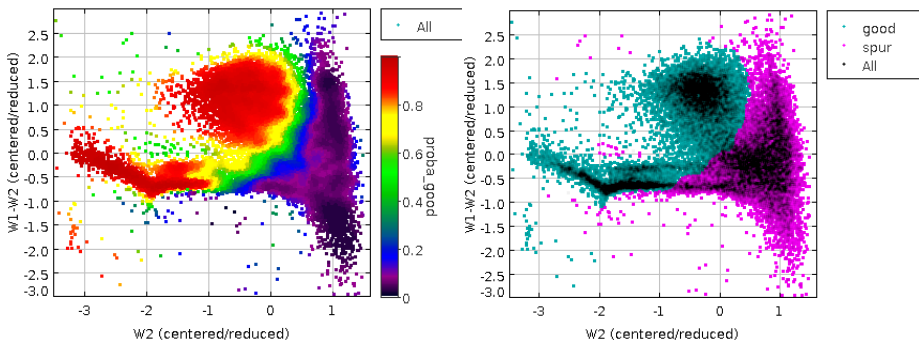
□ NWAY / ARCHES: photometry

- Likelihoods (distributions) computed by kernel smoothing (sample point estimator, $k=75$)
 - ▶ left: $p(\vec{m}|good)$
 - ▶ right: $p(\vec{m}|spur)$



□ NWAY / ARCHES: photometry

- Left: classification result $p(\text{good})$
- Right: binary classification Good/Spurious ($p(\text{good}) > 0.5$, $p(\text{good}) < 0.5$)



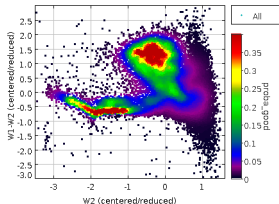
□ Merging with positional proba

- Accounting for photometric likelihoods (simplified Eq. 154 of Pineau et al. 2017)

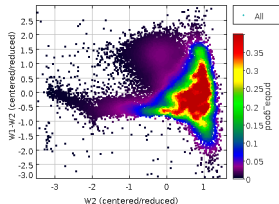
$$p(\text{real}|x, \vec{m}) = \frac{p(\text{real}|x)p(\vec{m}|\text{real})}{p(\text{real}|x)p(\vec{m}|\text{real}) + (1 - p(\text{real}|x))p(\vec{m}|\text{spur})}$$

$p(\text{real}|x)$ = purely positional probability

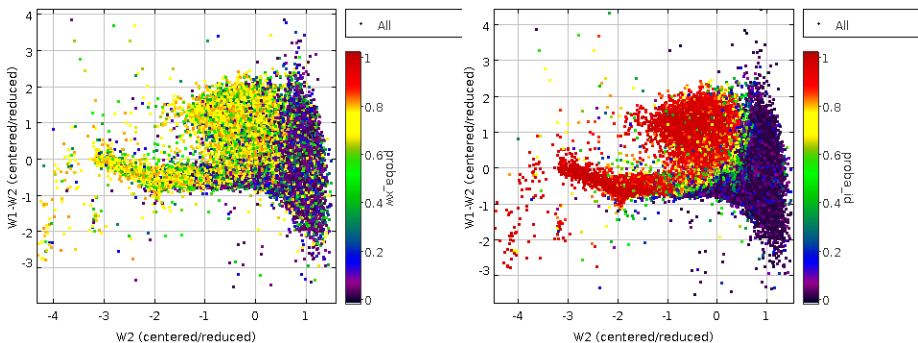
$p(\vec{m}|\text{real}) =$



$p(\vec{m}|\text{spur}) =$

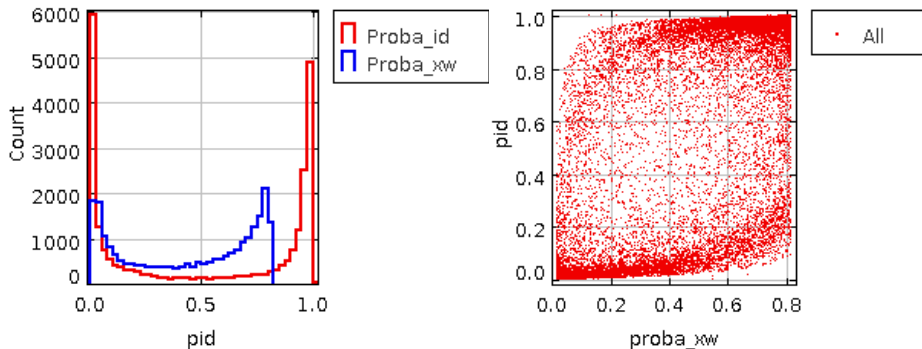


□ NWAY / ARCHES: photometry



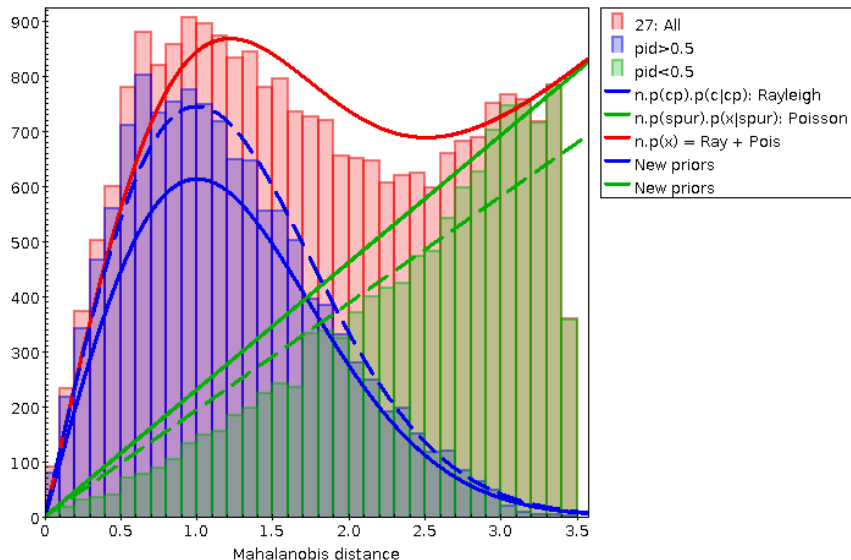
- Left: positional probabilities
- Right: probabilities accounting for photometric likelihoods

□ NWAY / ARCHES: proba id



- Clearer separation of low and high probabilities

□ NWAY / ARCHES: proba id



□ XMM vs SDSS DR8

Testing the same method to cross-match XMM and SDSS


XMM vs SDSS DR8

Choose tables to cross-match

IX/50/xmm3r6s X SDSS DR8

VizieR SIMBAD My store

[XMM-Newton Serendipitous Source Catalogue 3XMM-DR6 \(XMM-SSC, 2016\)](#)
468,440 rows






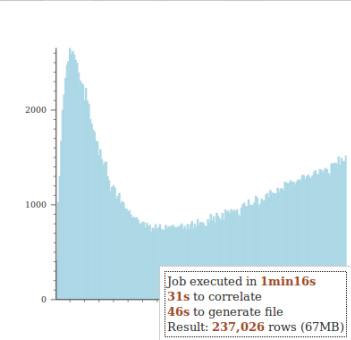
Show options

Begin the X-Match

Visualize and manage your cross-matching jobs

List of X-match jobs

| Table 1 | Table 2 | Options | | |
|---------------|----------|--|---------------------|---|
| IX/50/xmm3r6s | SDSS DR8 | fixed radius  radius: 10 arcsec area: All sky | 18/05/2018 at 22:31 | completed   |



Job executed in **1min16s**
31s to correlate
46s to generate file
Result: **237,026 rows (67MB)**

<http://cdsxmatch.u-strasbg.fr>

□ XMM vs SDSS DR8

- Quick-and-dirty test in 4 dimensions

- ▶ Simple 10 arcsec cross-match
- ▶ Keep only *primary unresolved* objects having a *clean photometry*
- ▶ Mahalanobis distance: $d_\sigma \approx \frac{d}{\sqrt{\left(\frac{SC_POSERR}{\sqrt{2}}\right)^2 + RA_ERR \times DE_ERR}}$
- ▶ Lazy learning samples definition:
 - ★ 19676 “real” associations: $d < 1''$ && $d_\sigma < 1.5$
 - ★ 7784 “spurious” associations: $d > 8''$ && $d_\sigma > 6$
- ▶ User defined prior $p(cp)$ going to $d_{\sigma,max} = 5$
- ▶ From Eq. 149 of Pineau et al. (2017):

$$p(cp|d_\sigma) = \frac{1}{1 + \frac{1-p(cp)}{p(cp)} \frac{2}{d_{\sigma,max}^2} e^{-\frac{d_\sigma^2}{2}}}$$

□ XMM vs SDSS DR8

- Using the CDS prototype service:

Put the data files into the distant server

```
./classif.bash put good xmm_sdss8.unres.good.csv # 19676 rows
```

```
./classif.bash put spur xmm_sdss8.unres.spur.csv # 7784 rows
```

```
./classif.bash put data xmm_sdss8.unres.all.csv
```

Performs the classification of the data and save the result

```
./classif.bash kdc samplepoint -k 25 -p good:0.55\;spur:0.45 -ho > result.csv
```

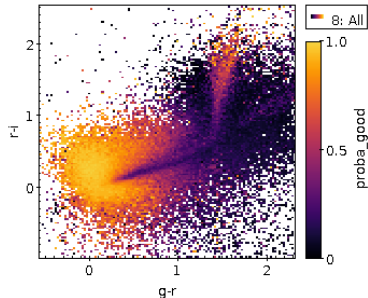
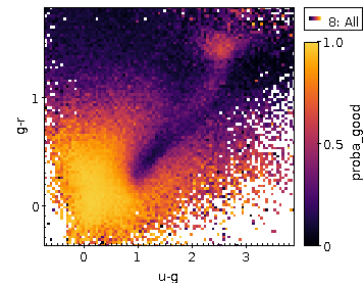
Ask for the confusion matrix by self-classifying the LS

```
./classif.bash kdc samplepoint -k 25 -p good:0.55\;spur:0.45 -cr
```

- Confusion matrix:

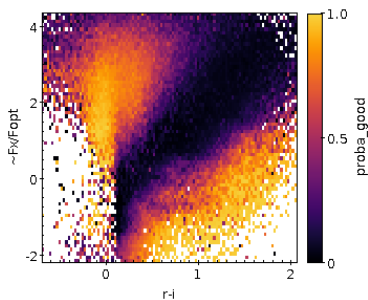
| actual \ predicted | good | spurious |
|--------------------|--------|----------|
| good | 86.91% | 13.09% |
| spurious | 12.28% | 87.72% |

XMM vs SDSS DR8



Mean of the 4D KDC output probabilities in

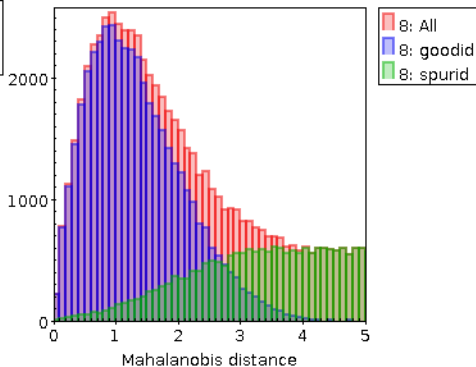
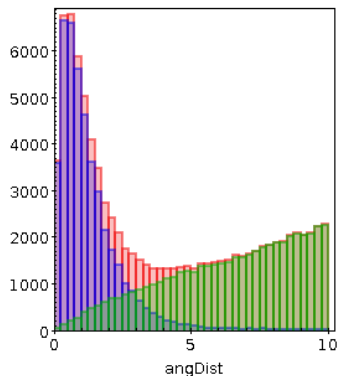
- $u - g$ vs $g - r$
- $g - r$ vs $r - i$
- $r - i$ vs $\propto F_X/F_r$



□ XMM DR7 vs SDSS DR8

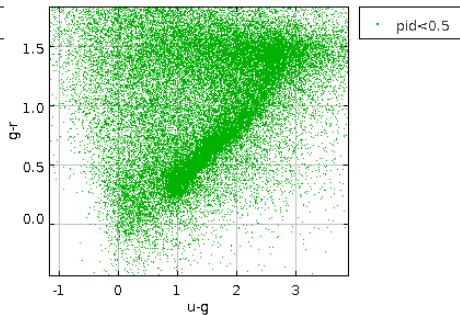
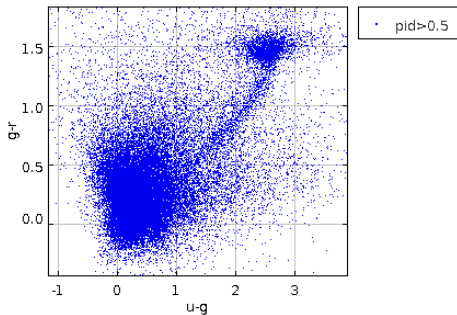
Using 4D photometric likelihoods to compute final proba $p(cp|d_\sigma, \vec{m})$, and considering:

- real matches as $p(cp|d_\sigma, \vec{m}) > 0.5$
- spurious matches as $p(cp|d_\sigma, \vec{m}) < 0.5$



XMM DR7 vs SDSS DR8

- $u - g$ vs $g - r$ diagrams of estimated as real and estimated as spurious associations.





Travail exploratoire

□ Index spatio-temporel

- Coordonées sphériques + axe temporel entremêlés
 - ▶ HEALPix + axe auxiliaire (Z-order curve 3D)
- But:
 - ▶ répondre rapidement à des requêtes spatio-temporelle
 - ▶ décrire des regions spatio-temporelles avec des MOC
- Implémenté: $(\alpha, \delta, t) \rightarrow idx, idx \leftarrow center(\alpha, \delta, t)$
- A implémenter: cone search + time range \rightarrow MOC
- A tester: taille des regions (MOC) en pratique

□ Perspective

Vers le Big Data (Spark)?

Test préliminaire:

- Cross-match Gaia DR1 (1.1 G srcs) vs IGSL3 (1.3 G srcs)
- 5 arcsec \rightsquigarrow 1.6 G associations
- Cluster de 9 machines reformées \rightsquigarrow 10 min
- Algorithme amélioré (kd-tree) à tester