Finding counterparts for all-sky X-ray surveys with N WAY : a Bayesian algorithm for cross-matching multiple catalogues

M. Salvato, J. Buchner, T. Budavári et al, 2018 MNRAS 473, 4937

09.06.2022

OUTLINE

- Matching objects from different catalogs
- N WAY program
- Specific example: maching a X-ray (2RXS, XMM Newton slew Survey) and infrared (AllWISE) catalogs

Cross - matching of catalogues.

Necessary for:

- Finding object for precise analysis.
- Population study of given type

Possible problems:

- Matching straightforward for catalogs with precise positions.
- Most objects in the catalogs of GW, GRB cannot be indentified.
- Large uncertainties in positions for gamma ray, X-ray, far infrared or some radio catalogs (idntification of quasars).
- Different catalogs: all-sky, but shallow; pencil-beams deep but small solid angle.

Object characterization

- position
- proper motion
- parallax
- radial velocity, redshift
- photometric data
- spectroscopic data

But without subarcsec precision of position source cannot be identified easily

Matching catalogs

- Likelihood Ratio (LR) method (Sutherland and Saunders 1992) Take into account coordinates (with relative errors), source number densities, and magnitude distribution of the sources. The method estimates the ratio between the likelihood that a given source from catalogue B is the correct counterpart to a source detected in catalogue A, and the likelihood of being a source in the background.
- N WAY – an algorithm based on two-steps Bayesian approach

NWAY algorithm: main features

https://github.com/JohannesBuchner/nway

- Matching of N catalogues simultaneously.
- Computation of all combinatorially possible matches
- Consideration of partial matches across catalogues, i.e. the absence of counterparts in some catalogues.
- Taking into account the positional uncertainties and the source number densities.
- Computation of the probability that there is no match.
- Computation of the probability of each possible match.
- Incorporating magnitude, colour or other information about the sources of interest, refining the match probabilities.

NWAY - main steps

- Finding combinatorially all possible matches.
- Computing each match probability from number densities, separation distances and positional errors alone, taking into account the chance of a random alignment.
- For each source of the primary catalogue compute (a) the probability that this source does not have a counterpart and (b), assuming this source has a counterpart, compute the relative probability for each possible match.
- Refining the probabilities by additional prior information.

All catalogs must contain position, positional error and the size of area covered by the catalog must be known.

Input and output catalogs

Input:

Primary Catalogue
x1
x2
...

2nd Catalogue
b1
b2
...

3rd Catalogue
c1
c2
...

Output:

Primary Cat. Entry	2nd Cat. Entry	3rd Cat. Entry	Proba-bility	
x1	b1	c1	...	
x1	b1	c2	...	
x1	b1	(none)	...	
x1	b2	c1	...	
x1	b2	c2	...	source x1
x1	b2	(none)	...	
x1	(none)	c1	...	
x1	(none)	c2	...	
x1	(none)	(none)	...	
x2	...	...	...	source x2

Matching with additional prior information

- Multiple priors can be used from any of the input catalogues. One can use any discriminating information (e.g. magnitudes, colours, morphology, variability, etc.).
- It is possible to input pre-constructed information or compute the distributions from the catalogues themselves based on secure distance-only matches

Probability for an individual association.

- We have k catalogs (i = 1,.., k)
- Our primary catalog has index i=1
- Let N_i denote number of entries in catalog "i"
- $\nu_i = N_i/\Omega_i$ - source surface density on the sky
- Probability of chance alingment on the sky of k physically unrelated objects P(H)

$$P(H) = N_1/\prod_{i=1}^{k} N_i = 1/\prod_{i=2}^{k} N_i = 1/\prod_{i=2}^{k} \nu_i \Omega_i$$

- To account for non-uniform coverage, P(H) is modified by a "prior completness factor" c (which gives the expected fraction of sources with reliable counterpart - for example for catalogs that do not cover the same area $\Omega_{i>1} \neq \Omega_1$)

$$P(H) = c/\prod_{i=2}^{k} \nu_i \Omega_i$$

Probability for an individual association.

- $P(D|H)$ is derived as:

$$P(D|H) = 2^{k-1} \frac{\prod \sigma_i^{-2}}{\sum \sigma_i^{-2}} \exp\left[-\frac{\sum_{i<j} \psi_{ij}^2 \sigma_j^{-2} \sigma_i^{-2}}{2 \sum \sigma_i^{-2}}\right]$$

ψ_{ij} pairwise angular separation between catalogs i, j
σ_j σ_j - position uncertainties

Probability for an individual association.

- Comparison of two hypotheses for a association: all sources are identified with the same object (H_1), sources are coincidentally aligned (H_0) (H_0 - no counterpart hypothesis).
- Posterior of hypotheses comparison

$$\frac{P(H_1|D)}{P(H_0|D)} \sim \frac{P(H_1)}{P(H_0)} \times \frac{P(D|H_1)}{P(D|H_0)}$$

$$B = \frac{P(D|H_1)}{P(D|H_0)}$$

$$P(H_1|D) = \left[1 + \frac{1 - P(H_1)}{B \cdot P(H_1)}\right]^{-1}$$

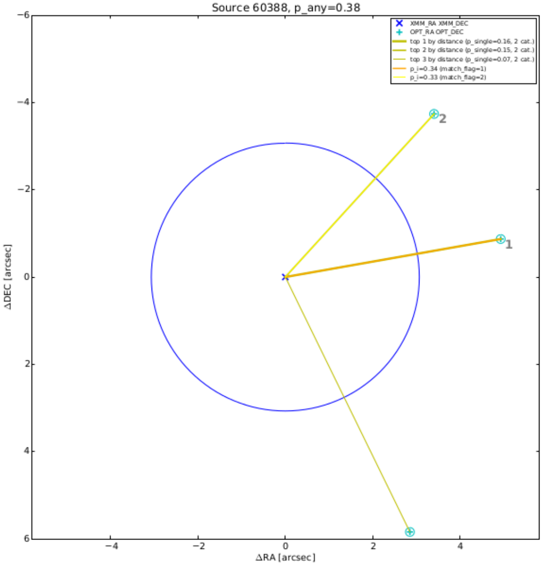
- For each entry in primary catalog the posteriors of all possible associations are normalized to unity and $P(H_0|D)$ is computed

Probability for an individual association. cont.

- $p_{any} = 1 - P(H_0|D)/\sum P(H_i|D)$
  p_{any} small - little evidence for association, most probable no-association case
  $p_{any} \approx 1$ - strong evidence for at least one of the association to another catalog
- Relative posterior probabilities of possible associations are renormalized with H_0 excluded.

$$p_i = P(H_i|D)/\sum_{i>0} P(H_i|D)$$

Example of small p_{any}



Taking into account additional information

- Specific classes of astronomical objects show distinct distribution on colour, magnitude or other parameters, compared with the field population distributions.
- Likekihood ratio coming from angular distance f(r) could be modified by a factor

$$LR = \frac{q(m)}{n(m)} \times f(r)$$

$q(m)$ and $n(m)$ are associated with the magnitude distributions of source (e.g. X-ray sources) and background objects, respectively, but additionally contain sky density contributions.

Additional information on magnitudes in Bayesian frame

- In Bayesian framework two likelihoods are combined by considering two independent observations: one for the positions D_ϕ and one for magnitudes D_m (for example).

$$P(D|H) = P(D_\phi|H) \times P(D_m|H) = P(D_\phi|H) \times \frac{\bar{q}(m)}{\bar{n}(m)}$$

  $\bar{q}(m)$ - probability that target (X-ray) source has magnitude m
  $\bar{n}(m)$ - probability that field source has magnitude m

- The modifing factor is renormalized to

$$P(D_m|H) = \frac{\bar{q}(m)}{\bar{n}(m)} / \int \frac{\bar{q}(m')}{\bar{n}(m)'} \bar{n}(m')dm'$$

  so when m is unknown $P(D|H) = P(D_\phi|H)$ (m is marginalized over its distribution in general population $\int P(D_m|H)\bar{n}(m')dm'$)

Generalized case

- We consider arbitrary number of photometric bands, each consisting of a magnitude measurement m and measurement uncertainty σ_m

$$P(D_m|H) = \prod \frac{\int_m \bar{q}(m)p(m|D_m)dm}{\int_m \bar{n}(m)p(m|D_m)dm}$$

$p(m|D_m)$ - Gaussian error distribution with mean m and standart deviation σ_m

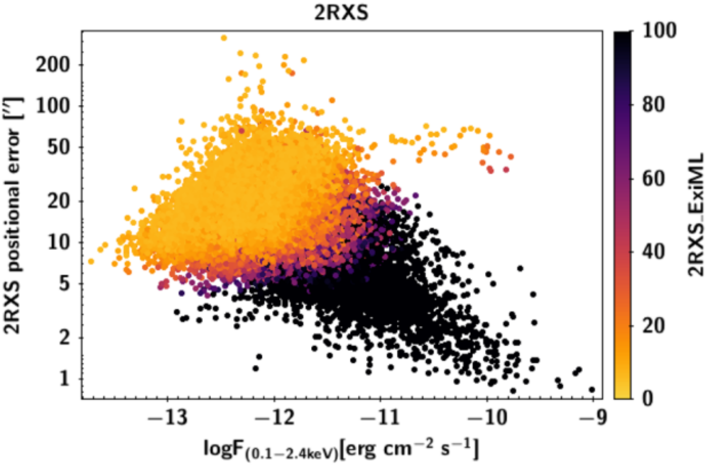
Autocalibration

- Distribution of $\bar{q}(m)$ and $\bar{n}(m)$ can be obtained by program working on deep field precise catalog using pure distance weighting
- Sources with high probability matches (for example dist_post > 0.9) contribute to $\bar{q}(m)$
- Sources ruled out as counterparts (for example dist_post < 0.01) contribute to $\bar{n}(m)$

ROSAT All sky X-ray survey

- Observations 1990-1991 in 0.1-2.4 keV
- Bright Source Catalogue (18916 sources)
- Faint Source Catalogue (105924 sources down to detection likelihood limit 6.5)
- 2RXS about 135 000 point - like sources, reanalysis of ROSAT data with new data reduction, and detection algorithms
- Auhors stude the area with $b > 15^o$, and at least 6^o and 3^o from LMC and SMC respectivelly.
- Sky coverage 30576 deg^2 with 106695 2RXS detections

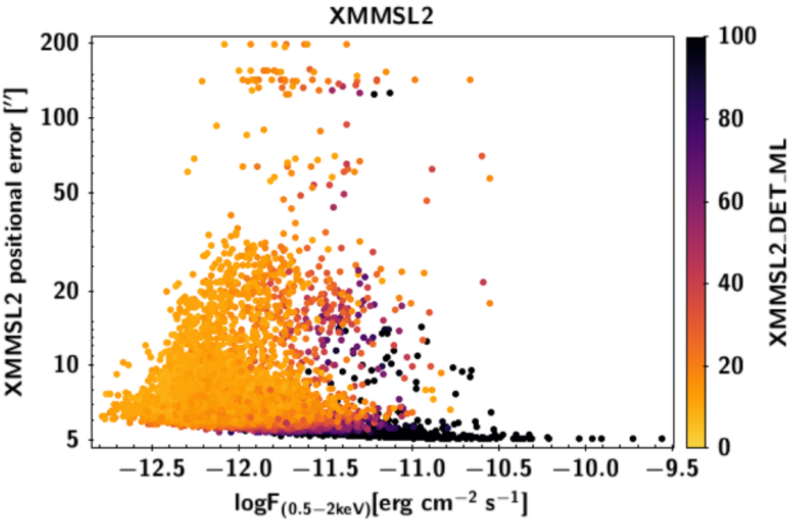
Position uncertainties and detection likelihood for 2RXS



XMM - Newton Slew 2 survey

- The XMM–Newton European Photon Imaging Camera pn (EPIC-pn) accumulates data during slews between pointed obser- vations
- 84 % of the sky
- In the area of the study there are 22306 X ray detections with at least 0.1 s of effective exposure
- detectors: 0.2-2keV, 2-12 keV and 0.2-12 keV

Position uncertainties and detection likelihood for XMMSL2



WISE and AllWISE catalog

- Wide-field Infrared Survey Explorer (WISE)
- Launched in 2009, all sky survey in 3.4, 4.6, 12 and 22 μ bands (W1, W2, W3, W4)
- Additional observations in W1 and W2.
- AllWISE data relase (2013, 747 million objects)

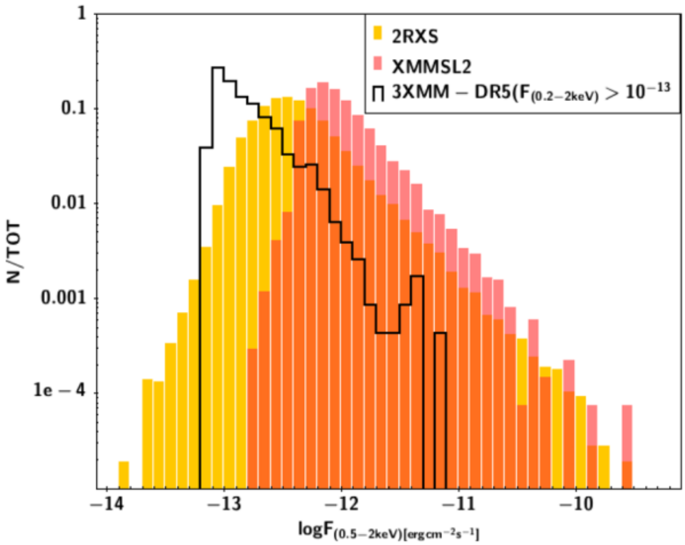


Figure 2. Flux distribution for the 2RXS (yellow), XMMSL2 (brown) and 3XMM-DR5 catalogues. The flux from the original bands has been transformed to the flux at (0.5–2 keV), assuming a Galactic $N_{\mathrm{H}} = 3(2.29)\mathrm{e}20\,\mathrm{cm}^{-2}$ and a power law of 1.7(2.4) for *XMM* SL2(2RXS) data,

AllWISE colour-magnitude prior

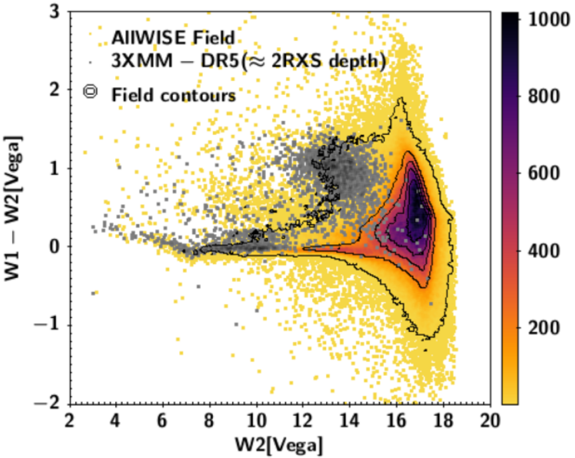


Figure 3. AllWISE colour-magnitude ([$W2$] versus [$W1 - W2$]) distribution of counterparts to the 3XMM-DR5 catalogue cut at the depth of 2RXS (grey) compared with the AllWISE distribution (contours and density map) of all sources within 2 arcmin of the 3XMM-DR5 sources.