

(EPIC) Source Detection & Catalogues

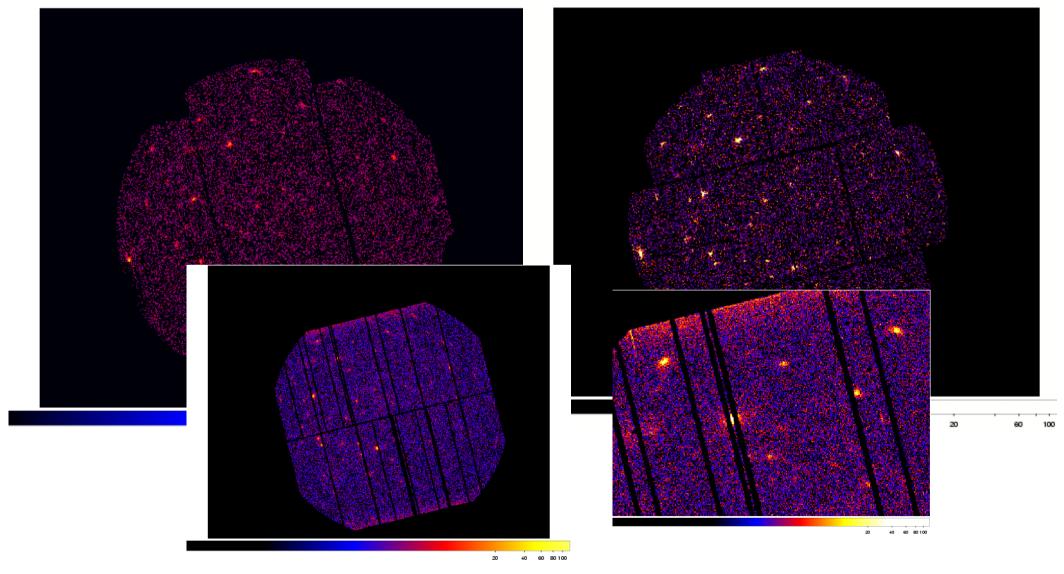
Aitor IBARRA & Carlos GABRIEL

XMM-Newton Science Operations Center - ESA (with main contributions from G. Lamer, AIP, and A. Read, UoL + SSC)

In the beginning there are images

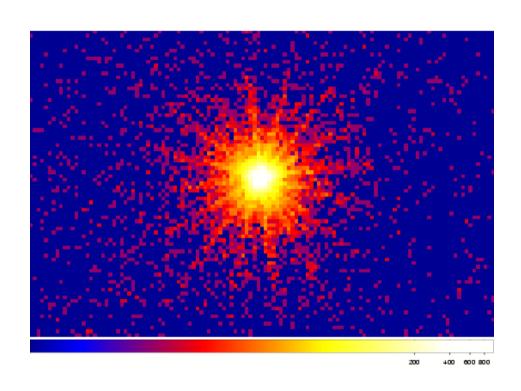


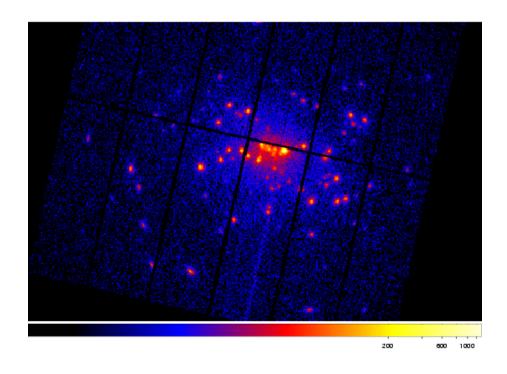
Image production from events lists = collapsing events onto X-Y plane



In the beginning there are images







Source searching



- >> source searching means basically looking for
- * significant fluctuations, which are
- * compatible with sky sources,
- * lying on top of *more* or *less* smooth distributions,
- * avoiding to get *fooled* by **detection defects**

>> goal:



to **maximize** source detection sensitivity **minimizing** number of fake detections

SAS source searching: preparatory steps



Looking for (small) fluctuations on top of distributions

- >> maximization of S/N ratio for sources to be found
- ... cleaning calibrated event lists against flaring periods
 - >> produce high energy background lightcurves
 - + define a threshold + produce GTIs

- ... taking into account the different source spectral characteristics
 - >> apply **band-passes** for deriving corresponding **images**

Source searching: preparatory steps



>> produce high energy background lightcurves

evselect table=PNevt:EVENTS

expression='#XMMEA_EP&&(PI>10000)&&(PATTERN==0)'

rateset=pn_back_lightc.fits timebinsize=10 withrateset=yes

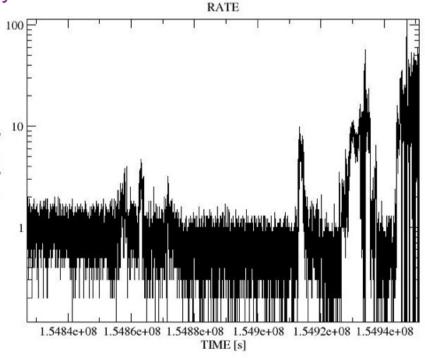
maketimecolumn=yes makeratecolumn=yes

+ define a threshold

dsplot table=pn_back_lightc.fits x=TIME y=RATE &

+ produce GTIs

tabgtigen table=pn_back_lightc.fits expression="RATE < 1.50" gtiset=pn_back_gti.fits

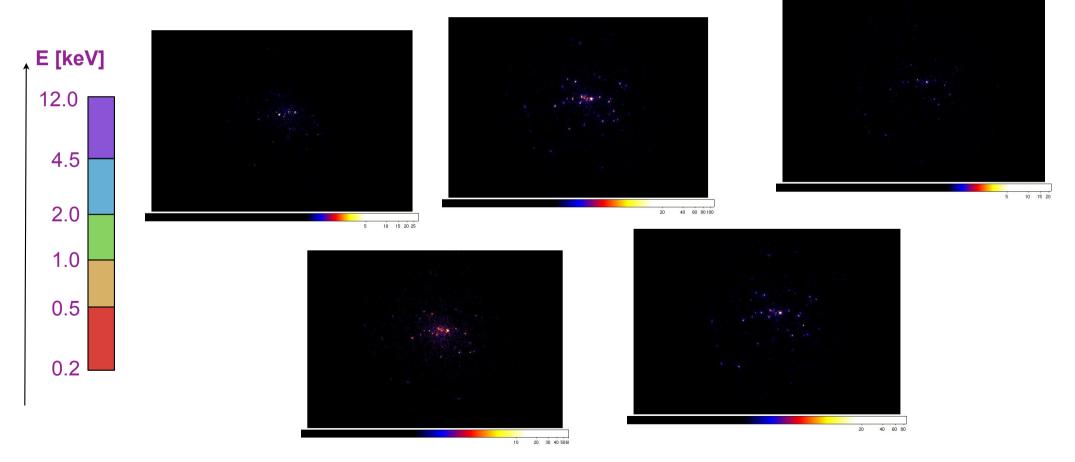


pn_back_lightc.fits

Source searching: preparatory steps

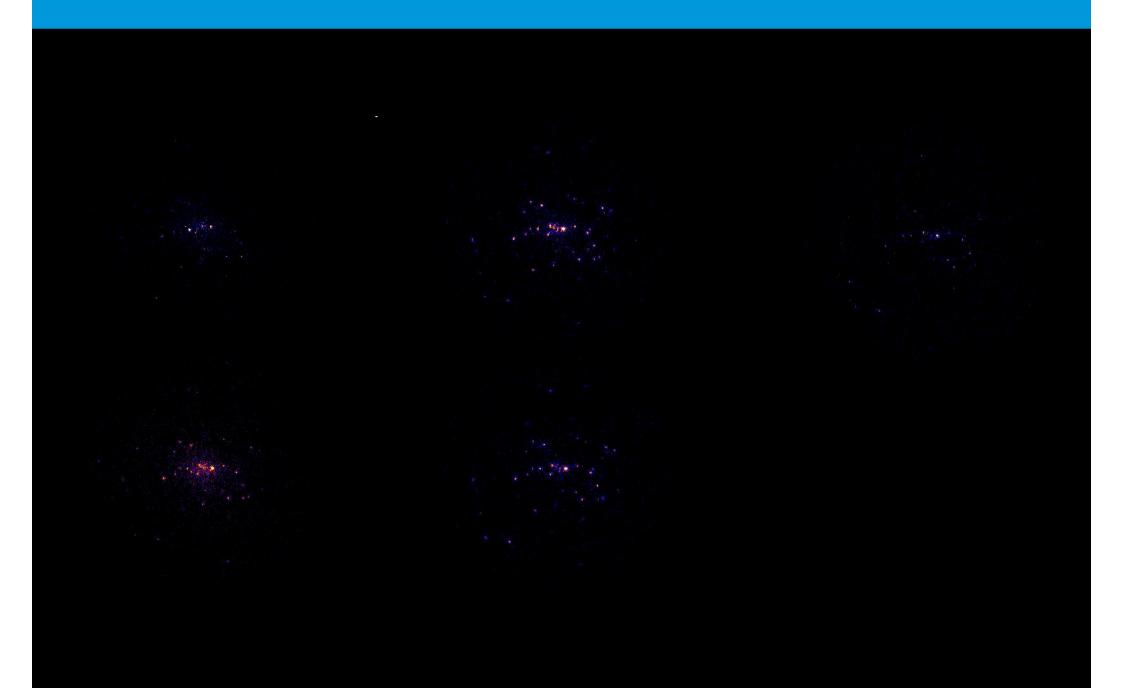


- >> apply band-passes for deriving corresponding images
- * selection of energy bands depend on the scientific purpose:
- basic XMM-Newton bands:



Source searching: preparatory steps





SAS source detection tasks



Two methods of performing source detection on EPIC datasets:

1) edetect_chain

perl script running
all these tasks
consecutively

lask	Purpose	input data sets	output data sets
eexpmap	creation of exposure maps	images, attitude files	exposure maps
emask	creation of detection masks	exposure map	detection mask
eboxdetect (local mode)	sliding box detection	images, exposure maps, detection mask	local box list
esplinemap	creation of background maps	images, exposure maps, detection mask, local box list	background map
eboxdetect (map mode)	box detection using bkg map	images, exposure maps, detection mask, background maps	map detect source
emldetect	maximum likelihood fitting	images, exposure maps, detection mask, background maps	final source list
esensmap	creation of sensitivity maps	exposure map, detection mask, background map	sensitivity map

import data aata

2) ewavelet mexican hat wavelet algorithm for detecting both point and extended sources.

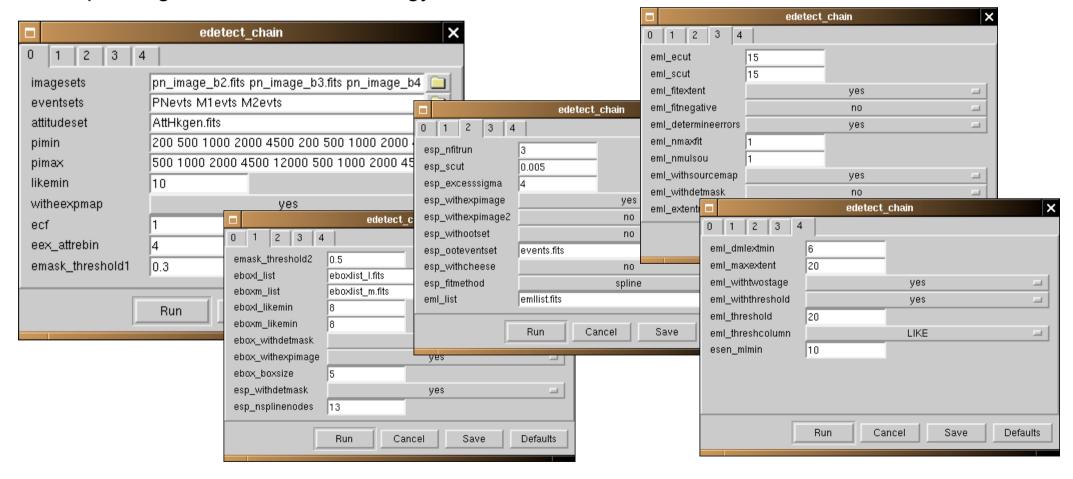
Easy to use and efficient, but less reliable source parameters than those from edetect_chain

edetect chain



edetect_chain is able to work on up to 240 images at one time

→ eg ("PPS approach"): running simultaneously the whole detection chain with 15 input images corresponding to the 5 standard energy bands of each EPIC camera



Source detection tasks 1 x 1



To quantify the significance of detected signals we need to know the observing conditions, eg., how long we have exposed the different parts of the detectors

>> effective exposure time for each detector point

spatial quantum efficiency, filter transmission, mirror vignetting and FOV calculating for each attitude bin the exposure values projected onto the sky

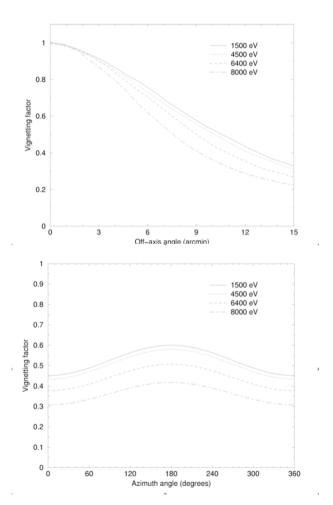
[data taken from Calibration Files]

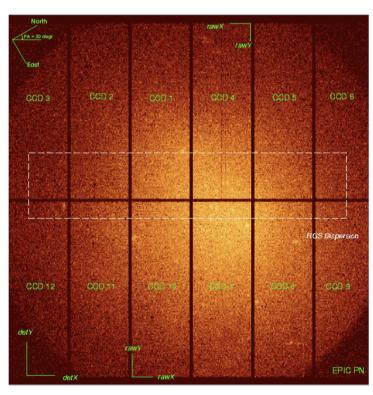
>> **instrument maps** containing **spatial efficiency** of the instrument (unit=[seconds])

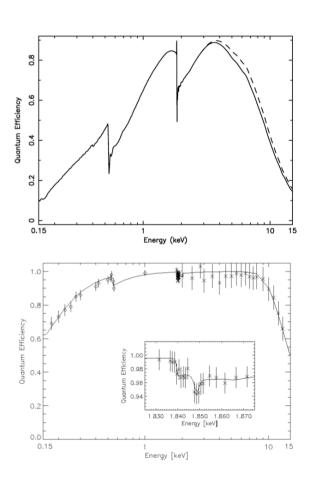
Source detection tasks



spatial quantum efficiency, filter transmission, mirror vignetting and FOV calculating for each attitude bin the exposure values projected onto the sky







edetect_chain: eexpmap



(PN) example of single task commands issued:

1) create the multiband exposure maps by eexpmap:

- event energy is assumed to correspond to PI channel boundaries given (parameter pimin and pimax)
- Bad pixels as listed in the bad pixel extension of the input file are excluded from the instrument maps.
- Depending on flag selections in the image, the surroundings of bad pixels and border pixels are also excluded from the instrument maps.
- output in detector or sky coordinates (same as input image)

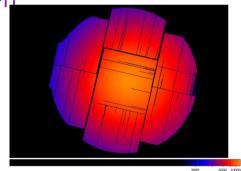
eexpmap attitudeset=xxxATTxxx.FIT eventset=xxxEVLlxxx.FIT imageset=xxxIMAGExxx.FIT pimin="200 500 1000 2000 4500" pimax="500 1000 2000 4500 12000" expimageset="pn_1exp.fits pn_2exp.fits pn_3exp.fits pn_4exp.fits pn_5exp.fits"

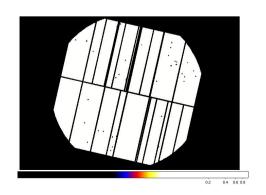
2) create the detection maps (area defined which is suitable for source detection):

emask expimageset=pn_2000.fits threshold1=0.5 detmaskset=**pn_mask.fits**

Main criterium - valid area is only **area_i** such that:

exposure_i > threshold1 * MAX(exposure)





finding source candidates



Source detection is performed by eboxdetect

3) sliding box detection (local mode):

eboxdetect in "local mode"

a) **source counts** accumulated in 3x3 or **5x5** pixel window (parameter boxsize) **background** from surrounding 7x7 or **9x9** pixels (40 or 56 pixels respectively)

Detection of extended sources doubling up to 3 times the pixel size in consecutive runs

Background subtracted source counts calculated applying correction factors to account for respective fractions of source counts falling on source and background area

 \mathbf{n} = detection box size

Enboxed energy fractions in source / background box: $\alpha = \sum_{n \times n} PSF$ / $\beta = \sum_{(n+4)\times(n+4)} PSF - \sum_{n \times n} PSF$

Raw box counts & raw background counts: $\mathbf{C} = \sum_{n \times n} i mage$ & $\mathbf{Bg} = \left(\sum_{(n+4)\times(n+4)} i mage - \sum_{n \times n} i mage\right) / ((n+4)^2 - n^2)$

PSF corrected and background subtracted counts: $SC = C - Bg * n^2 / [\alpha - \beta * n^2 / ((n+4)^2 - n^2)]$

sliding box detection / I - cont.



Detection likelihoods given as:

 $L = -\ln p$ p = probability of Poissonian random fluctuation of background counts in cell resulting in \geq observed source counts $(p \text{ using incomplete Gamma function } \Gamma(a,x) \text{ as function of raw source and raw background counts in the detection box)}$

In case of simultaneous detection over several bands, likelihoods are added (!) and transformed into equivalent single band detection likelihoods

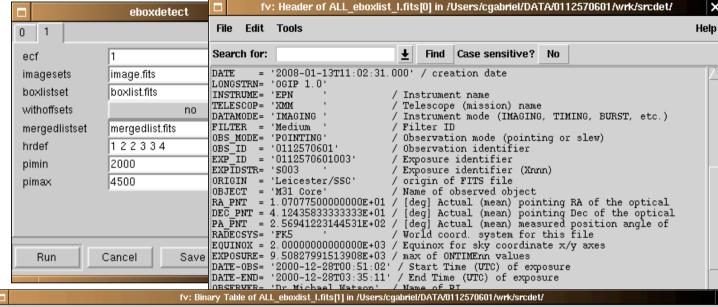
$$L = \Gamma(n_{\text{band}}, \sum_{i=1,n} L_i)$$

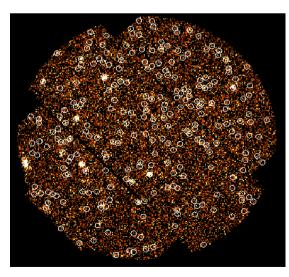
eboxdetect usemap=no likemin=10 withdetmask=yes detmasksets=pn_mask.fits \
 imagesets="PNIM_1000.FIT PNIM_2000.FIT PNIM_3000.FIT PNIM_4000.FIT PNIM_5000.FIT" \
 expimagesets="pn_1000.fits pn_2000.fits pn_3000.fits pn_4000.fits pn_5000.fits" \
 pimin="200 500 1000 2000 4500" pimax="500 1000 2000 4500 12000" \
 boxlistset=eboxlist local.fits

eboxdetect > FITS tables









	□ BOX_ID_SRC	☐ ID_INST	_ ID_BAND	□ SCTS	☐ SCTS_ERR	□ BOX_CTS	□ X_IMA	X_IMA_ERR	□ Y_IMA
elect	J	J	J	E	E	E	E	E	E
_ All				counts	counts	counts	image pixels	image pixels	image pixels
Invert	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify
1	1	0	0	8.606420E+02	7.444390E+01	2.980000E+02	7.676847E+02	8.919184E-03	3.484796E+02
2	1	1	0	4.769174E+02	5.399276E+01	1.610000E+02	7.676847E+02	8.919184E-03	3.484796E+02
3	1	1	1	1.047846E+02	2.559736E+01	3.500000E+01	7.676847E+02	8.919184E-03	3.484796E+02
4	1	1	2	1.673634E+02	3.275788E+01	5.700000E+01	7.676847E+02	8.919184E-03	3.484796E+02
5	1	1	3	1.310721E+02	2.913927E+01	4.500000E+01	7.676847E+02	8.919184E-03	3.484796E+02
6	1	1	4	5.990199E+01	1.374246E+01	1.900000E+01	7.676847E+02	8.919184E-03	3.484796E+02
7	1	1	5	1.379532E+01	1.220497E+01	5.000000E+00	7.676847E+02	8.919184E-03	3.484796E+02
8	1	2	0	2.138778E+02	3.677129E+01	7.100000E+01	7.676847E+02	8.919184E-03	3.484796E+02
9	1	2	1	1.882268E+00	1.038807E+01	4.000000E+00	7.676847E+02	8.919184E-03	3.484796E+02
10	1	2	2	8.139063E+01	1.627813E+01	2.500000E+01	7.676847E+02	8.919184E-03	3.484796E+02
11	1	2	3	7.403590E+01	2.290094E+01	2.400000E+01	7.676847E+02	8.919184E-03	3.484796E+02

creating background maps



Modelling the background is a key issue for source detection

EPIC background has three main components

a) Photons:

- astrophysical, dominated by thermal emission at lower energies (unresolved cosmological sources)
- solar wind charge exchange
- single reflections from out of FOV, out-of-time events, etc

b) Particles:

- soft proton flares
- internal (cosmic-ray induced) background, direct (on CCDs) / indirect (fluorescence of S/C)

c) electronic noise

- bright pixels, columns, etc, readout noise, etc.

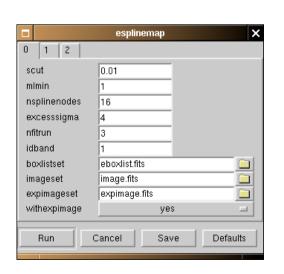
>> complex issue - so far in source detection by default phenomenologically:
2D-spline

creating bkg maps: esplinemap



- 4) creation of background maps (done per detector and band)
- a) Cutting-out sources (using source brightness dependent radius), esplinemap blanks out areas with sources detected by eboxdetect >> cheesed image
- b) n x n (default=12) spline fits >> smoothed background map for entire image

esplinemap bkgimageset=pnback_1000.fits scut=0.005 imageset=PNIM_1000.FIT nsplinenodes=12 \ withdetmask=yes detmaskset=pn_mask.fits withexpimage=yes, expimageset=pn_1000.fits \ boxlistset=eboxlist local.fits



0.05 0.1 0.15 0.2 0.25 0.3 0.3E pnback 2000.fits

pnback 5000.fits

European Space Agency

eboxdetect in map mode



5) box detection using the background maps

in **map mode** background is taken from background maps determined by **esplinemap** >> improved detection sensitivity compared to local detection map

```
eboxdetect usemap=yes likemin=8 withdetmask=yes detmasksets=pn_mask.fits \
    imagesets=... expimagesets=... pimin=... pimax=... bkgimagesets="pnback_1000.fits ..." \
    boxlistset=eboxlist_map.fit
```

Output table:

- one raw per input image for each detected source (source table)
- number of summary rows containing broad band results for each EPIC telescope
- combined results for all EPIC telescopes taken together

Source table:

- count rates and source positions including statistical errors + fluxes + ...
- detection likelihoods (per band and total) given for each source I = -In p with p=probability of random fluctuation of counts resulting in $N_{Counts} \ge C_{obs}$
- if several images inputted then hardness ratios are calculated:

$$HR_i = (B_m - B_n) / (B_m + B_n)$$

max likelihood fit: emldetect



6) maximum likelihood fitting for getting final source list

Method: **Simultaneous maximum likelihood PSF fit** to source count distribution (convolved with a **source extent model**) in all bands with following free parameters:

- * source location (α, δ)
- * source extent (gaussian sigma)
- * source count rates in each band

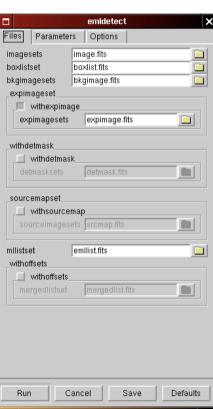
| constrained to same best-fit value | in all energy bands of each EPIC

I individual best-fit value in each band

Second loop for fitting two PSFs to "extended sources" - if better > recalculation

PSF fitting may be performed in **single source** or in **multi-source** mode. In multi-source mode sources with overlapping PSFs are fitted simultaneously (up to 6)

Energy conversion factors (ECF) supplied for conversion of count rates into correct flux values. The ECFs depend on filter and pattern selection



emldetect & extended sources



PSF fitting by emldetect is most crucial step for characterization of extended sources Default (also for 3XMM) is the convolution of PSF with a β model for source extent

$$f(x,y) = \left(1 + \frac{(x-x_0)^2 + (y-y_0)^2}{r_c^2}\right)^{-3\beta+1/2}$$
 β = 2/3 (canonical for surface distribution of clusters)

Fitting procedure minimizes the C - statistic

 $C = 2 \sum (e_i - n_i \ln e_i)$ e= expected model & n= number of counts in pixel i

Detection likelihood for each input image IM is $L_{IM} = C_{NULL} - C_{BEST}$ C_{NULL} is C of Null-hyp.

Detection likelihood obeys L = -ln(P) with P probability that source is spurious Extended likelihood L_{ext} calculated in analogy with C_{NULL} = best fitting point source model

Second fitting loop against source confusion: 2 source models simultaneously fitted (only for brighter sources)

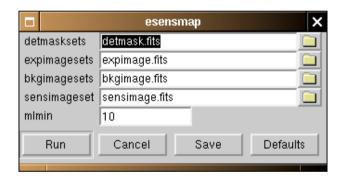
sensitivity maps: esensmap

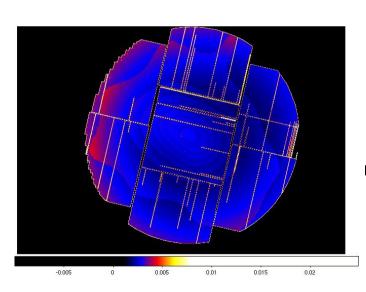


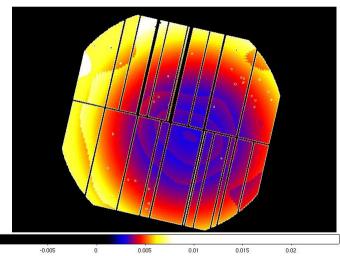
7) creation of sensitivity maps (called for each detector and band)

esensmap *expimagesets=* pn_1000.fit *bkgimagesets=*pnback_1000.fits *detmasksets=*pn_mask.fits \ *mlmin=*10. *sensimageset=*sens map1.fits

Sensitivity map == point source **detection upper limits** (vignetting corrected source count rate corresponding to the likelihood of detection as specified in the parameter file) for each image pixel.







position rectification: eposcorr / catcorr



off-edetect_chain) position rectification using optical catalogues

correlation with optical source catalogue, checking whether there are offsets in RA and DEC which optimize the correlation

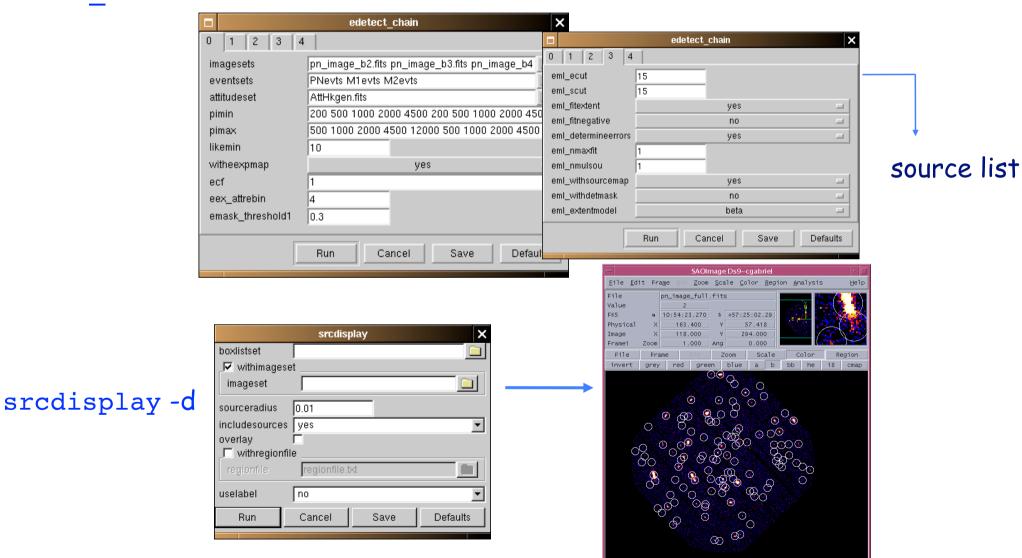
Used in 4XMM-PPS:

catcorr srclistset=emllist.fits catset=catextract.ds mingood=10 minfit=5 maxoffset=10
using not only USNO-B, but also 2MASS and SDSS >> covering 85% of all observations

displaying sources: srcdisplay



edetect chain -d



Full reprocessing + 4XMM-DR12



- ESAC finished in mid-2020 the reprocessing of all the XMM-Newton data on behalf of the Survey Science Centre (SSC) @ IRAP (Toulouse) —> 4XMM-DR10
- Uniform archive in terms of processing and calibration
- Since then, 4XMM-DR11 took place in September 2021

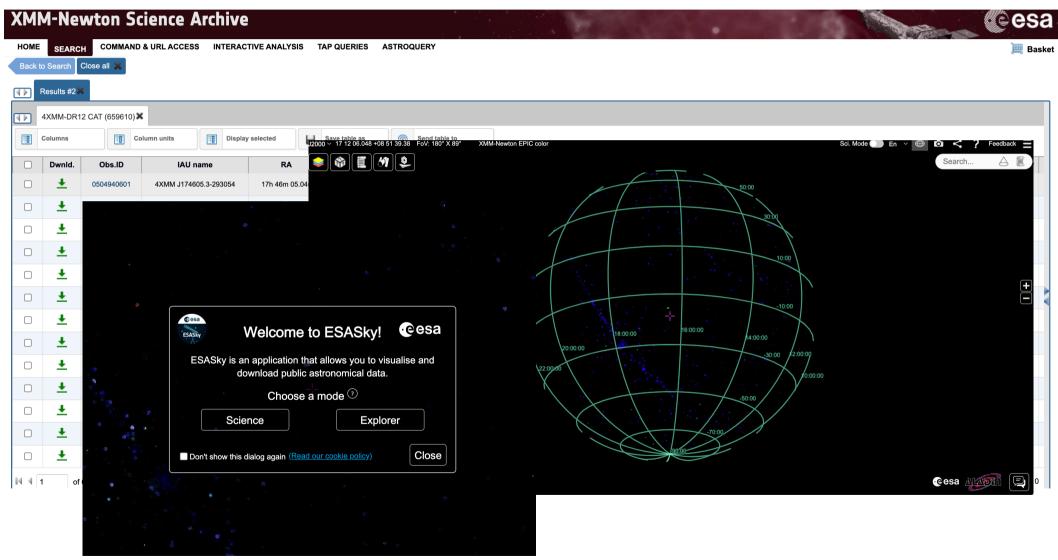
4XMM-DR12 catalogue = largest catalogue of X-ray sources [Webb et al., 2020]

	4XMM-DR12	4XMM-DR11	Increment
Number of observations	12712	12210	502
Number of 'clean' observations (i.e., observation class < 3)	10543	10247	296
Observing interval	03-Feb-00 31-Dec-21	03-Feb-00 17-Dec-20	1 yr
Sky coverage, taking overlaps into account (≥ 1ksec exposure)	1283 sq.deg	1239 sq.deg	44 sq.deg
Number of detections	939270	895415	43855
Number of 'clean' detections (i.e., summary flag < 3)	821953	787963	33990
Number of unique sources	630347	602543	27804
Number of 'cleanest' (summary flag = 0, not in high-background fields) extended detections	20169	19077	1092
Number of detections with spectra	337058	319565	17493
Number of detections with timeseries	336776	319292	17484
Number of detections where probability of timeseries being constant is $< 1 \times 10^{-5}$	7697	7283	415

- $(0.2 12 \text{ keV}) \longrightarrow \sim 2.3 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$
- $(0.2 2 \text{ keV}) -> \sim 5.2 \times 10^{-15}$
- $(2 12 \text{ keV}) \longrightarrow \sim 1.2 \times 10^{-14}$.

Full reprocessing + 4XMM-DR14



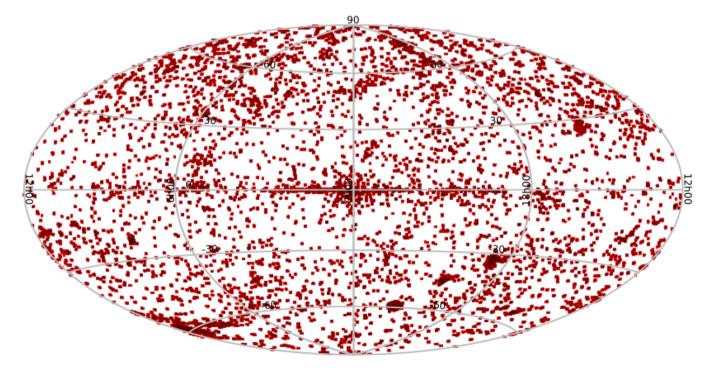


https://sky.esa.int/

4XMM-DR12 properties

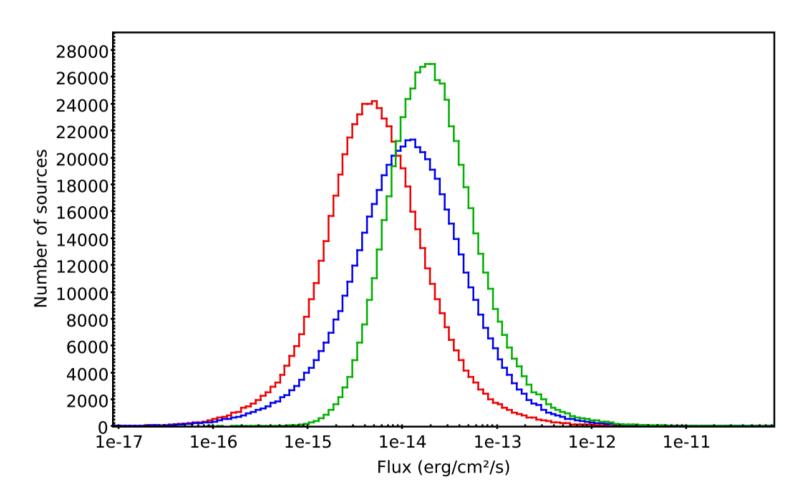


- The 4XMM-DR12 catalogue is about 5% larger than the 4XMM-DR11 catalogue.
- In terms of the number of X-ray sources, combining the 4XMM-DR12 and 4XMM-DR12s catalogues gives a catalogue that is similar in size to the Chandra Source Catalogue but 4XMM-DR12+4XMM-DR12s contain twice as many individual sources as the Chandra catalogue



4XMM properties - detection limits



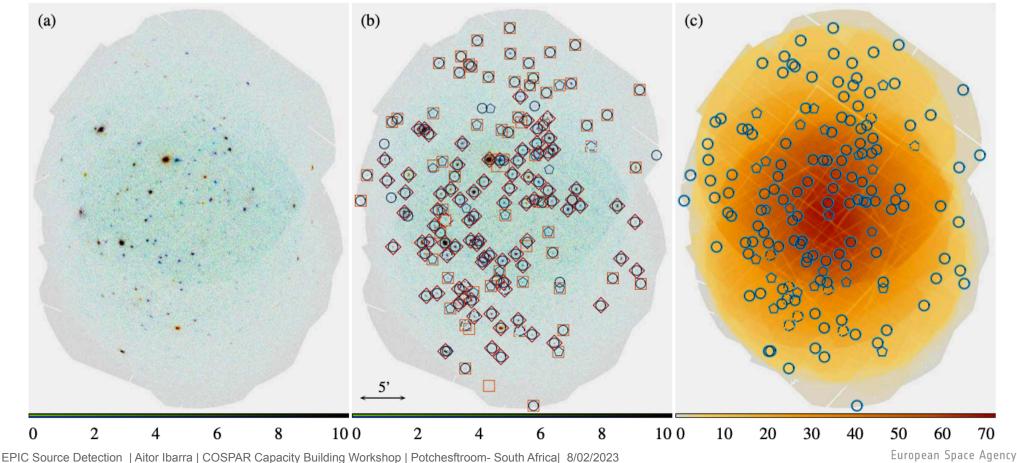


0.2-2.0 keV: Red2.0-12.0 keV: BlueTotal Band: Green

4XMM-DR12 - Stacked Catalogue



- A second independent stacked catalogue is compiled in parallel by the XMM-Newton SSC, called 4XMM-DR9s, where the letter 's' stands for stacked [Traulsen et al., 2020]
- This catalogue lists source detection results on overlapping XMM-Newton observations.
- The construction of the first version of such a catalogue, 3XMM-DR7s, is described in <u>Traulsen et al.</u> (2019).



4XMM-DR12 - Stacked Catalogue

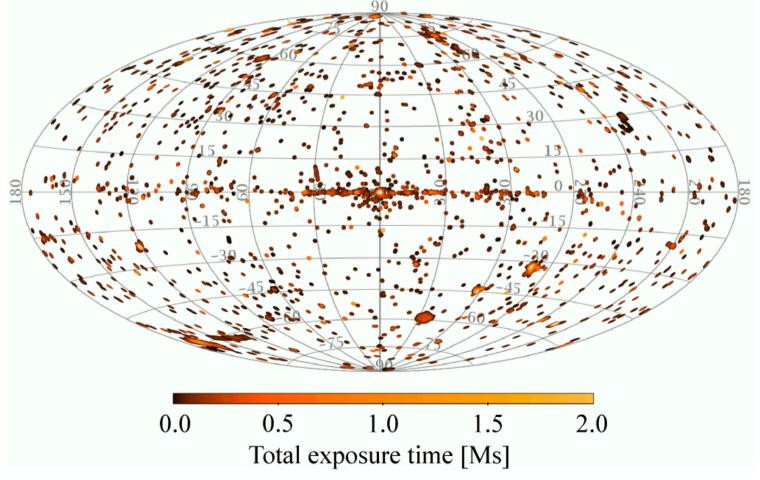


Most of the stacks are composed of 2 observations and the largest has 372.

• The catalogue contains 386043 sources, of which 298626 have several contributing observations.

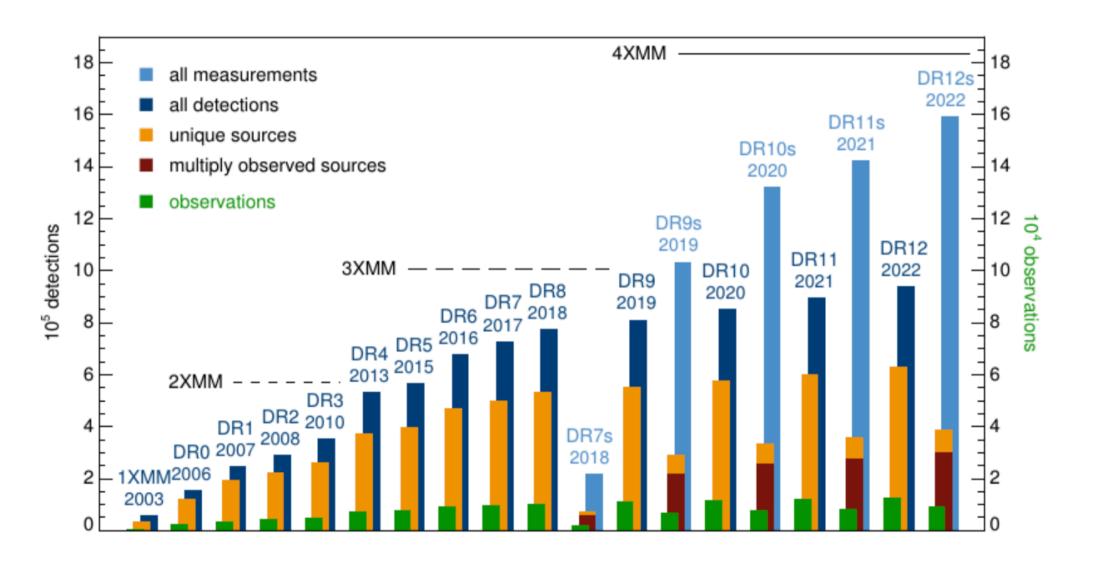
Stacking observations allows yet fainter sources to be detected in sky regions observed more

than once,



XMM-Newton Catalogue: in numbers esa





Detecting sources in overlapping fields @esa



Main example for overlapping fields is the Mosaic Mode

Basic definitions:

- series of stable pointings with EPIC cameras in FF / Window mode, continuously taking data (same filter)
- only 1 PN offset map, taken by first pointing
- angular offset between pointings within [0.2 - 60] arcmin
- shortest integration time per pointing = 1500 sec
- whole observation included in one ODF - if observation not possible within one revolution, then several obs's.

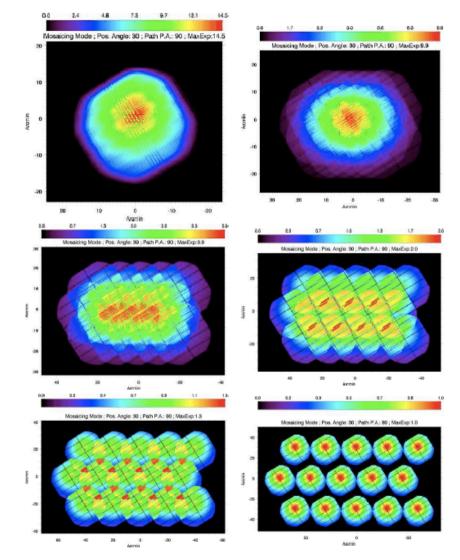


Figure 1: These figures simulate the EPIC-pn effective exposure maps achieved for a mosaic consisting in 5x3 individual pointings and for different angular offsets (1.5, 10, 15, 20 and 30 arcmin). The duration of a single pointing is taking as unit for these exposure maps. The relative position angle of the instruments has been arbitrarily set to 30degrees.

Simulated exposure maps of a 5x3 mosaic taken for angular offsets of 1.5', 10', 15', 20' and 30' with flat exposure per pointing, ignoring slews.

From P. Rodriguez TN

Mosaic mode: analysis



Question to SAS / PPS:

- how to treat this data?
- definitions:
 - separate data corresponding to different pointings as if they were different exposures (ignoring slews)
 - treat them coherently for source detection, eg. one call to eboxdetect (map mode) and emidetect

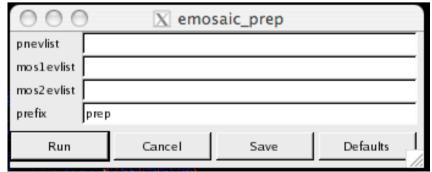
Our PPS scheme: separate mosaic ODF into n single pointing ODFs >> process them "normally"

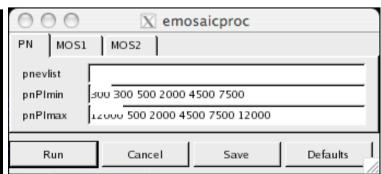
Our SAS scheme:

- >> normal reduction to large single event file (epicproc)
- >> separation of events from different pointings
 - >> one event file per point per instrument through emosaic_prep
 - >> coherent source detection of (overlapping) chosen fields through emosaicproc

Source detection working with all EPIC data (memory ~ map size can become an issue)

>> GUIs

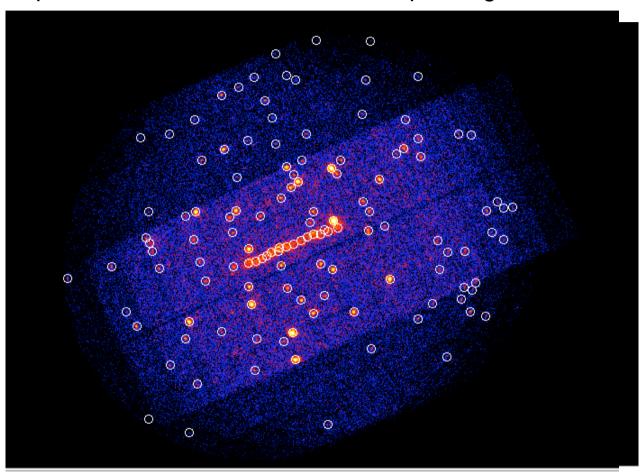




Mosaic mode - results - one mosaic



Jupiter observation 0200080701 - 4 pointings:



Problems with mosaic mode data:

- > number of pointings can be large
 >> enormous needs of memory
- > combination of not overlapping data not necessary at all
- >> decision about which points to be combined and how to do recombination of source lists left to the observer

Remember - right combination:

 more efficient source detection instead of WRONG source detection (wrong LHs if no separation)

Detection using 3 spectral bands for each pointing and instrument, [400-1000], [1000-2000] and [2000-10000] eV

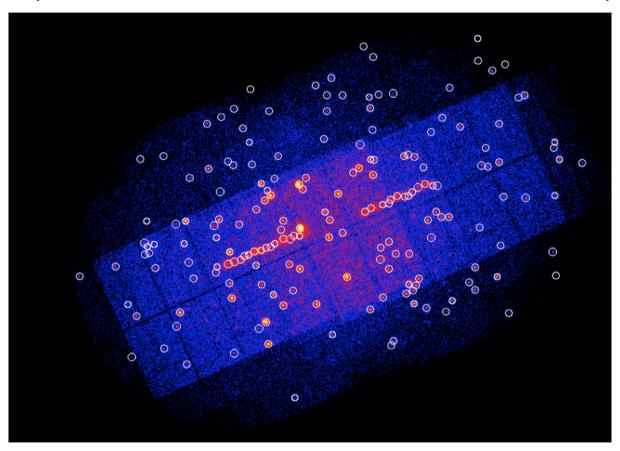
>> 36 images combined for eboxdetect (map mode) and emidetect

>> so far no PPS implementation

Mosaic mode - results - combining ODFs



Jupiter observation 0200080201+ 0200080701 - 4+4 pointings:



>> can be used by any overlapping observations, also with single pointing obs's taken at different times

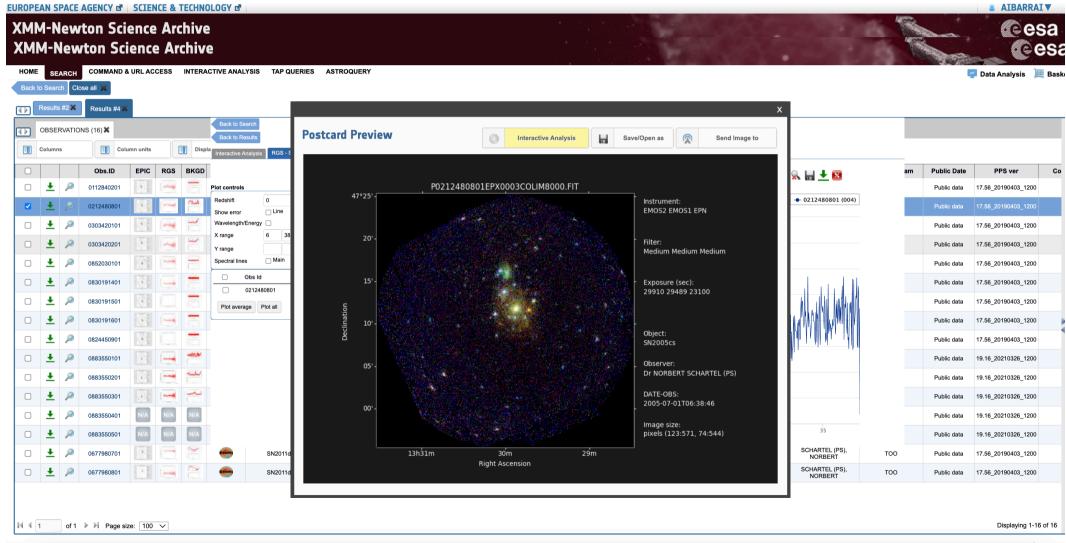
Also using 3 spectral bands for each pointing and instrument, [400-1000], [1000-2000] and [2000-10000] eV

>> 72 images combined for eboxdetect (map mode) and emldetect

How to get XMM-Newton data? The XSA



https://www.cosmos.esa.int/web/xmm-newton/xsa



COPYRIGHT © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED. V15.1.2-OPER (17-11-202

How to get XMM-Newton data? The XSA



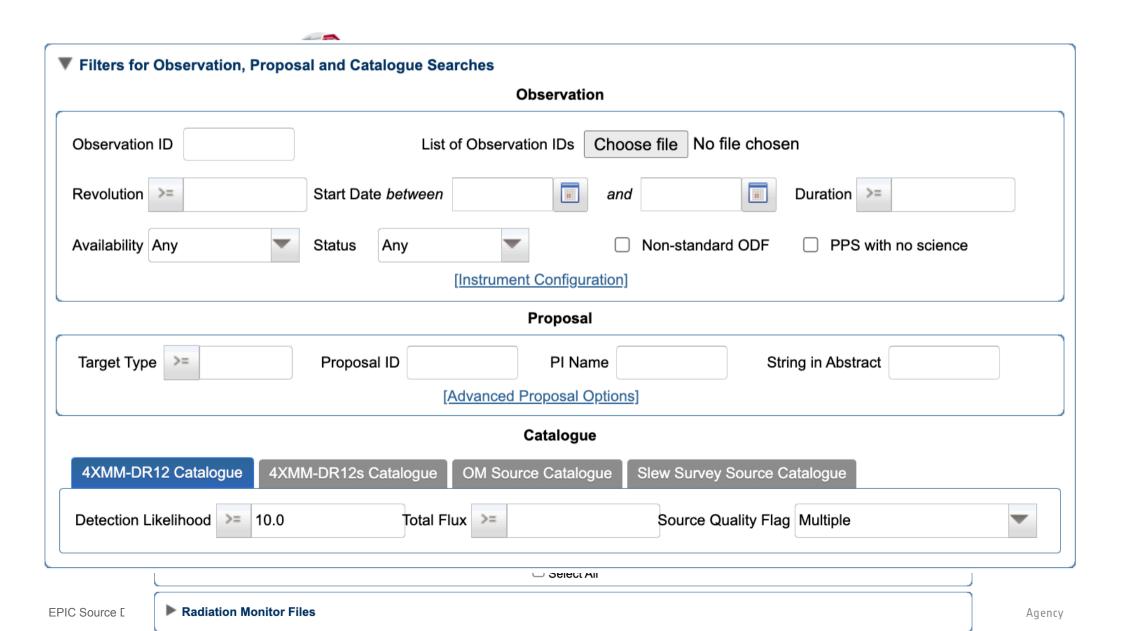
https://www.cosmos.esa.int/web/xmm-newton/xsa



1. INTRODUCTION The module astroquery.esa.xmm_newton is a python interface for querying the XMM-Newton Science Archive (XSA) web service. 2. USAGE Once python is installed you can do: pip install --pre astroquery or if you have a previous astroquery version installed: pip install --pre --upgrade astroquery More details can be found in astroquery instructions: https://astroquery.readthedocs.io/en/latest/ 3. EXAMPLES See: https://astroquery.readthedocs.io/en/latest/esa/xmm_newton.html 2 1. Getting XMM-Newton data: >>> from astroquery.esa.xmm_newton import XMMNewton >>> XMMNewton.download data('0505720401',level="PPS",extension="PDF",instname="M1",filename="result0505720401.tar") INFO: File result0505720401.tar downloaded to current directory [astroquery.esa.xmm newton.core] This will download all PPS files for the observation '0505720401' and instrument MOS1, with 'PDF' extension and it will store them in a tar called 'result0505720401.tar'. The parameters available are detailed in the API. For more details of the parameters check the section 3.4 at:

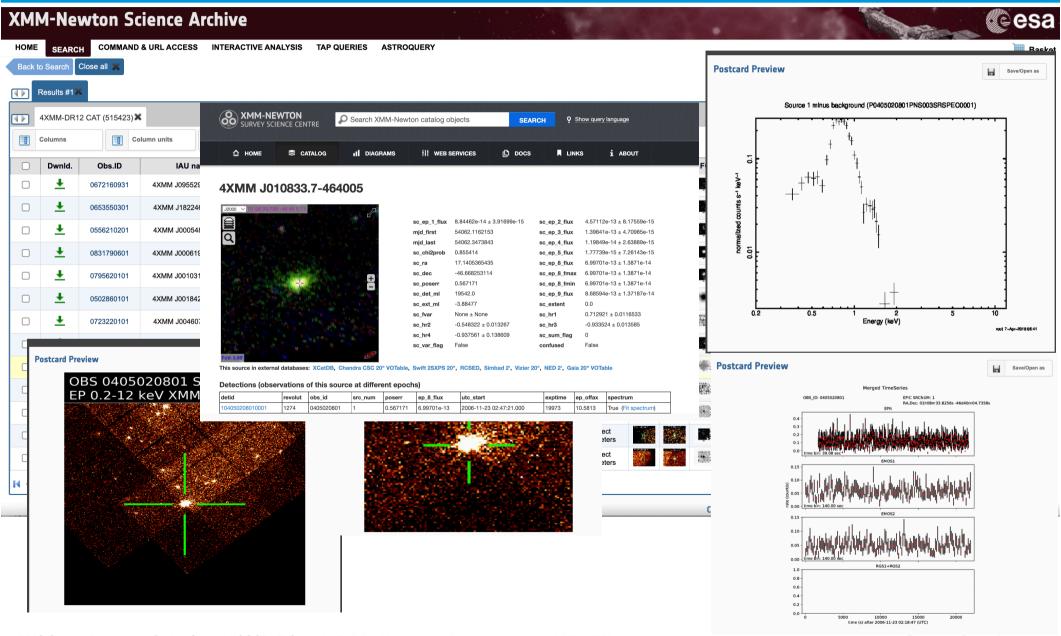
How to get 4XMM data? The XSA





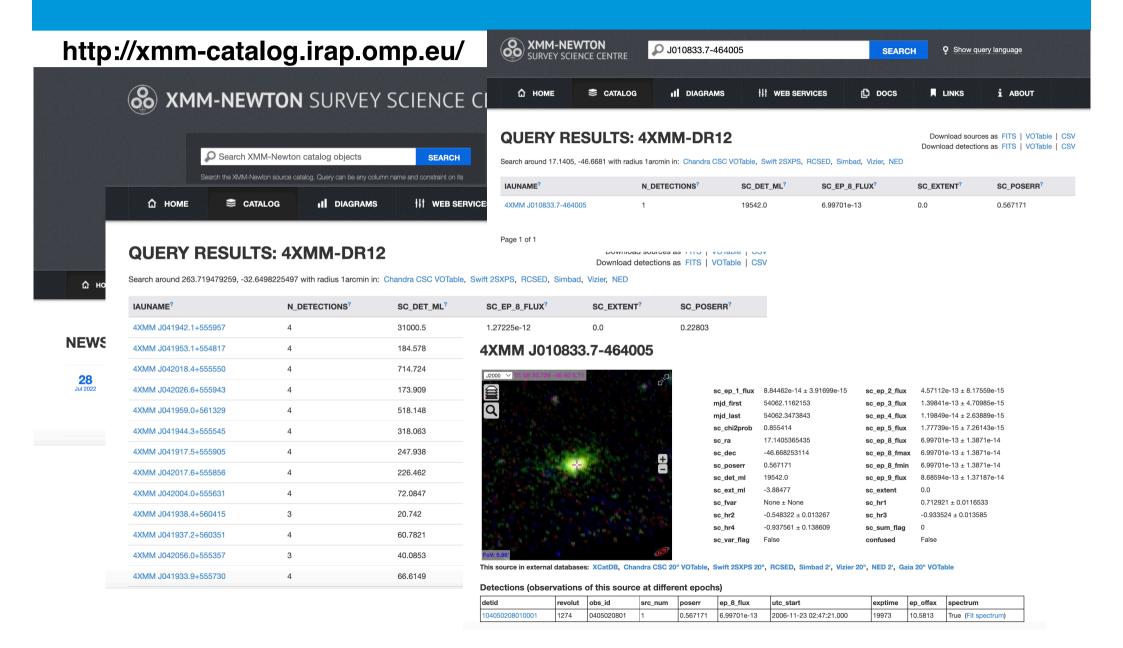
All the individual sources detected





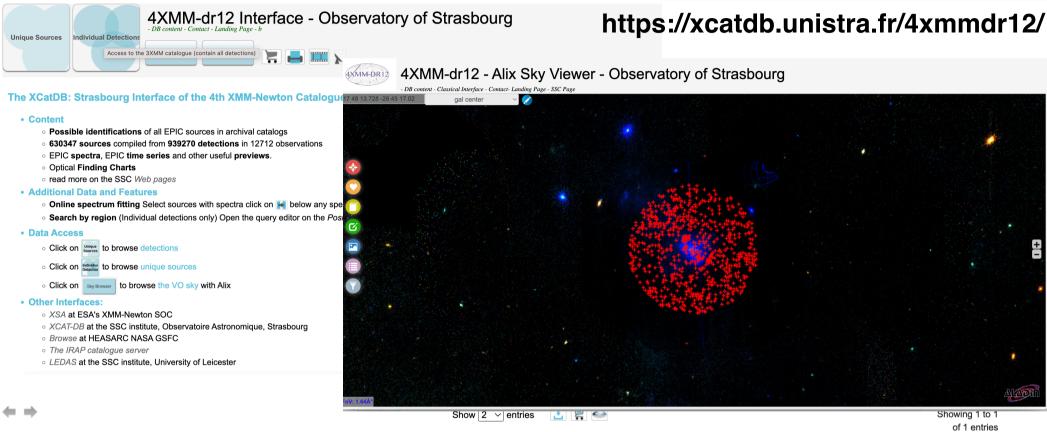
How to get (elsewhere) 4XMM data?





How to get (elsewhere) 4XMM data? XCatDB





Selection of Unique Sources

