ARCHES

The enhanced 3XMM catalogue (3XMMe)

S. Rosen
29 May 2015

1. Overview

The enhanced 3XMM catalogue (designated 3XMMe) is one of the core elements of the Arches project, representing the X-ray source basis for the cross-correlations with other multi-wavelength catalogues.

The 3XMMe catalogue is a derivative of the latest increment of the 3XMM catalogue, i.e. 3XMM-DR5, that was publicly released in April 2015 (Rosen et al., 2015 (submitted to A&A); http://arxiv.org/abs/1504.07051), tailored for the purposes of the Arches project. The 3XMM-DR5 catalogue, available from the XMM-Newton Science Archive (http://xmm.esac.esa.int/xsa/), the XMM-Newton Survey Science Centre (SSC) (http://xmmssc.irap.omp.eu) and other sites listed on the SSC site, contains 565962 detections arising from 396910 unique sources drawn from 7781 XMM observations. These numbers include detections from 356 sub-pointings made in mosaic mode.

The construction of the enhanced 3XMM catalogue involves 3 main elements

1. The cleaning (ejection) of detections that are considered to be of lower scientific reliability or quality.
2. Application of criteria to identify detections that are considered potential candidates for the 3 main science themes (i.e. active galactic nuclei (AGN), clusters of galaxies and the galactic plane) of the Arches project. The constituent detections of unique sources on the sky are examined to decide which unique sources, overall, meet the criteria to be assigned to each science theme.

3. Addition of other information that augments the scientific value of the catalogue.

The primary publicly released version of the catalogue is based on unique sources rather than separate detections. However, a detection-based catalogue will also be made available.

This document describes the production of the catalogue. Section 2.1 explains how cleaning of the catalogue was conducted. The application of filtering/flagging of detections/sources for science themes is described in section 2.2 while section 2.3 covers the flagging of detections/sources that lie in fields where contamination by prominent objects may affect the data in X-rays and other wavebands. Section 2.4 summarises the outcomes of the filterings in terms of object numbers. Section 3) outlines the additional information being included into the catalogue. Section 4 provides an overview of the content of the catalogue. The testing and validation of the enhanced 3XMM catalogue will be reported via work-package 8.

This version of the document (v2.1) relates to v2.0 of the enhanced 3XMM catalogue that is derived from the 3XMM-DR5 catalogue.

2. Cleaning and filtering

2.1 Cleaning

By cleaning we mean the removal, from the enhanced 3XMM catalogue, of detections that are deemed to be of lower reliability or quality. These may include likely spurious detections, for example, associated with previously unflagged 'hotspots' but also sources whose parameterisation is suspect, due, for example, to poorly defined background or location in complex regions, or which arise from fields that have other known issues. The following criteria have been adopted for the cleaning stage.

1. **Bad mosaic fields**: The 3XMM-DR5 catalogue contains 356 observations that were observed in mosaic-mode (see http://xmmssc.irap.omp.eu). However, 19 of these observations, containing 975 detections, are affected by issues with the ODF data (see watchout details under 3XMM-DR5 at the same website) and are removed from the 3XMMe catalogue.

2. **High background fields**: A subset of 568 XMM-Newton fields containing 21779 detections display high background levels, identified through visual screening - the identification process is somewhat subjective and has been applied rather conservatively. These are a result of the optimised flare filtering process in the pipeline processing used in the creation of 3XMM-DR5 that does not exclude cases where the background is persistently high throughout an observation, rather than just displaying discrete flaring during the observation. We have excluded all such fields and their detections because they tend to have low sensitivity and/or contain numerous spurious detections.
3. **Hotspots**: A number of detections in the 3XMM-DR5 catalogue that were neither automatically flagged up by the emldetect SAS task during pipeline processing, nor recognized during manual screening of the data (and thus received a clean summary flag (SUM_FLAG=0) during the creation of the catalogue), appeared to concentrate in localised features when mapped to the separate instrument detector coordinate systems. We refer to these, here, as 'hotspots', though their actual cause has yet to be fully established – some appear to be associated with occurrences of bright pixels that were not recognised during pipeline processing or subsequent visual screening. For a given instrument, to isolate the affected detections, all 3XMM-DR5 detections with a band 8 detection log-likelihood (L) > 6 in that instrument, that also had $0 \leq L < 6$ in other instruments and a quality flag (SUM_FLAG) ≤ 2, were mapped to detector coordinates for that instrument. Rectangular 'hotspot' regions were manually defined in detector coordinates that enclosed regions identified, visually, as being over-dense in detections. Detections with non-zero log-likelihoods in a given instrument lying within one of its hotspot regions were isolated and the unique set from all 3 instruments were then removed from 3XMMe. A total of 22397 unique 'hotspot' detections have been excluded from the enhanced 3XMM catalogue. This is conservative in that no attempt is made to retain detections that are seen in multiple instruments – for example, around 3400 detections that lie inside a detector hotspot region are detected with a log likelihood > 20 in each of the 3 instruments.
4. **Offaxis > 12 arcmins**: It was decided that to minimise the impact of edge effects from the different instruments and to avoid any issues with the more complex PSF profiles in the outer regions of the field of view, detections beyond 12 arcmins should be excluded. The offaxis location of the astrometrically corrected position of each detection was defined with respect to the

spacecraft boresight pointing position for the field. The reference to the spacecraft boresight location was made because it simplifies the later creation of masks. A total of 130119 detections were excluded as a result of this offaxis filter.

5. **Exposure < 5ks**: Detections from observations in which the exposure times (as defined by the <inst>_TEXP information at http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR5/3xmmdr5_obslist.html) of all active EPIC instruments were < 5000s were excluded, where <inst> is pn, M1 or M2. This was done to remove shallow (low sensitivity) fields. The use of the TEXP values was made because this represents the maximum exposure time of any CCD in the field as a whole, whereas the <inst>_ONTIME values in the 3XMM-DR5 catalogue represent the maximum exposure of the CCD in which a given detection occurs. This approach simplifies the downstream creation of masks. A total of 11991 detections from 669 observations were removed by this filtering.
6. **Rectification failure**: Fields where astrometric rectification of the field could not be performed (POSCOROK=N) were excluded. This is because such fields have less well defined astrometry systematics (on a per field basis) and the peak of the distribution of total (centroiding and systematic) position errors for detections from such fields is ~0.5" higher than for corrected fields. A total of 63888 detections from 1249 observations are excluded on this basis.
7. **Zero exposure**: A set of 207 detections have been isolated, which were found to have quality flag (SUM_FLAG)=0 (i.e. clean) but an EPIC ontime (EP_ONTIME) = 0. These are cases where the object is detected only by the presence of its wings because the object is essentially centred in a chip gap in each active instrument or because it lies on the very edge of the field of view. While potentially real, they are of highly suspect quality and are removed.
8. **MOS Window modes**: It was also decided to exclude fields where the pn was off and both MOS1 and MOS2 were off or being used in window modes (i.e. in any of Small Window, Large Window, Refresh-Frame-Store (RFS), Fast compressed/uncompressed). This was done to minimise complications in determining the background object densities of external catalogues when using the cross-correlation tool with XMM fields that contain significant blank sky area. A total of 45961 detections are in such fields.
9. **Observation class**: A decision was made to exclude all detections from fields with OBS_CLASS ≥ 4 to ensure use of the cleanest fields and to minimise problems for the multi-wavelength

catalogue cross-correlation step, as outlined in point 8. Such OBS_CLASS ≥ 4 fields have at least 10% of their sky area identified as potentially problematic. There are 112705 detections affected.

10. **Extension limit**: In characterising source spatial extent, the source detection and parameterisation process imposes an upper limit of 80" on the extent that a source can have. This causes a small number of extended sources to have their extent values clamped to this upper limit. As the extent values are thus not reliable, 3147 such cases are eliminated from 3XMMe.

Combining these elements, we find that a total of 280757 unique detections are excluded from the enhanced 3XMM catalogue compared to the base 3XMM-DR5 catalogue, leaving 285205 usable detections. These remaining detections are associated with 219788 unique sources and are drawn from 4802 XMM-Newton observations.

It should be emphasized that although almost half the detections in 3XMM-DR5 are removed from 3XMMe as a result of this cleaning process, most of the rejected detections are perfectly valid detections. The purpose of this cleaning has been to maximise the quality and accuracy of the detections (and unique sources) for the Arches project.

2.2 Science filtering and flagging

In addition to cleaning, detections/sources are identified as being suitable candidates for use in the specific science themes of the Arches project.

Each science team has established a set of criteria to be applied to **detections** to filter and flag them for their work and to provide indicators of suitability to external users. These are outlined below.

2.2.1 AGN

The following criteria are applied to detections

i. The galactic latitude, b_{II}, must satisfy $|b_{II}| > 20$ degrees. In fact the criterion applies not just to the specific detection but to all detections in the field in which it lies. In other words a detection will only be considered if the whole field meets the criterion. This is to simplify the later creation of masks and/or computation of sky area for fields that straddle the $|b_{II}| = 20$ boundary. In practice, the

test applied is that the absolute value of the galactic latitude of the field boresight be > 20.30 degrees, i.e. a field radius of 0.30 degrees (18 arcmins) is adopted.

ii. SUM_FLAG < 3. This effectively excludes all detections that were manually flagged/masked during 3XMM-DR5 catalogue screening.

iii. Detections are not in specifically identified fields. A list of 720 observations has been created that identifies a) fields that have been the subject of dedicated studies or b) overlap with large, well known objects (e.g. M31, Virgo Cluster) that complicate analysis. The list of sky areas avoided is provided in appendix A.

iv. Detections do not lie in fields affected by optically bright stars or galaxies (see section 2.3).

Note that there is some potential overlap between points iii and iv but each also has its own separate relevance in the filtering process.

2.2.2 Clusters

i. 0 < EP_EXTENT < 80 arcseconds. This only considers detections with real extent that is below the upper limit of 80 arcsecs imposed in the source detection step within the standard XMM-Newton pipeline processing.

ii. EP_EXTENT_ERR < 10. Excludes poorly constrained extent values.

iii. The galactic latitude must satisfy the same constraint ($|b_{II}| > 20.3$ degrees) as for AGN (2.2.1 (i)).

iv. EP_9_DET_ML > 10. Demands a minimum detection likelihood value of 10 in band 9 (XID band = 0.5-4.5 keV).

v. SUM_FLAG < 2. Excludes manually flagged detections and also detections with sum_flag=2, which the emldetect task itself, identifies as potentially suspect – generally the latter are detections that are extended and close to other sources or within the envelopes of other extended sources.

vi. 4 < offaxis < 12 arcmins. The offaxis angle for the detection is measured, in arcmins, from the spacecraft boresight for the observation. The lower boundary is intended to mitigate against the inclusion of extended target objects.

vii. At least one of the 3 EPIC instruments (<inst>=pn, M1, M2) has filter (<inst>_FILTER) = Thin1 and is being used in a full-frame mode, i.e. pn is in PrimeFullWindow or PrimeFullWindowExtended mode or MOS (1 or 2) is in PrimeFullWindow mode.

viii. EP_ONTIME > 5000. Requires that the local CCD containing the detection has > 5000s of exposure.

2.2.3 Galactic

No filtering/flagging criteria are specified.

2.2.4 Unique source flagging

Detections are flagged as eligible for use in the science themes according to the criteria outlined in sections 2.2.1 – 2.2.3. A unique source, however, is a combination of ≥ 1 constituent detections, reflecting a group of one or more detections that the catalogue creation software considers to be associated with the same single (unique) object on the sky. For unique sources comprising > 1 detection, the science flag assigned to the unique source (i.e. whether the unique source is associated with a science theme) depends on what decisions are taken about the validity of its constituent detections for the science. In such cases, some of its detections may pass the above tests while others fail.

Several scenarios are possible for identifying unique sources that are suitable candidates for the science themes. For example, flagging the unique source for a given science theme

I. if any (i.e. ≥ 1) constituent detection passes the relevant above criteria for that science.
II. only if ALL constituent detections pass the criteria (conservative approach)
III. if a majority (> 50%) of constituent detections pass the criteria.

In the current version (2.0) of the enhanced 3XMM catalogue we provide 3 separate flags for both the AGN and Cluster science themes. Two of the flags represent the first two options while the third flag is a numerical value that reflects the number of constituent detections in a unique source that pass the criteria, expressed as a fraction of the total number of

constituent detections. These flags give the user greater flexibility in selecting objects suitable for their science goals.

2.3 Bright stars and galaxies

Another issue affecting the quality of the data is the presence of bright stars or prominent galaxies whose presence may contaminate or complicate the analysis of data from other wavebands (e.g. in the optical or IR). To identify such cases, XMM field information was cross-matched against the Uppsala General Catalogue of Galaxies (UGC – e.g. http://cdsarc.u-strasbg.fr/viz-bin/Cat?VII/26D) for the galaxies and the Tycho-2 (e.g. see http://cdsarc.u-strasbg.fr/viz-bin/Cat?I/259) catalogue for the stars. For the galaxies, fields were noted where the galaxy, defined by its elliptical parameterisation in the UGC catalogue, overlapped > 20% of the area of the central 12 arc-minutes of the XMM field of view. For stars, any field containing a star brighter than magnitude B_T=6.7 (or V_T=6.7 if the B band value was not available) within 12 arcmins of the field centre, or B_T (or V_T) < 5.8, if within 15 arcmins of the centre, were noted – these limits were established empirically from visual inspection. The process of identifying the relevant fields involved an automated analysis but was backed up by visual inspection of fields. For example, fields containing stars with B_T <8 within 15 arcmins of the centre were also inspected, with some suspected problem cases also being noted, mainly where diffuse optical emission was present.

A total of 1211 unique fields containing problematic galaxies and stars were considered unsuitable for use within the AGN science theme. For the enhanced 3XMM catalogue, detections from any such ‘contaminated’ field were identified and considered unsuitable for the AGN science theme.

At the unique source level, three further flags are provided. These identify sources where any of the constituent detections are in affected fields, where all constituent detections are in affected fields, and also the fraction of constituent detections in affected fields.

2.4 Candidate detections and sources for AGN and Clusters.

As a result of applying the science filters, after the cleaning stage, the numbers of detections and unique sources considered suitable for the

AGN and Cluster science themes are shown in tables 1 and 2 respectively. The first column in each table indicates the number of detections that were assigned a flag indicating their suitability for the science goals. The second column indicates the number of unique sources in which at least 1 detection is flagged as suitable while the third column indicate the number of unique sources in which all detections are considered suitable. In many cases, a unique source comprises only one detection.

Table 1

AGN_FLAG (dets)	SC_AGN_ANY (src)	SC_AGN_ALL (src)
144801	123211	122264

Table 2

CLUS_FLAG (dets)	SC_CLUS_ANY (src)	SC_CLUS_ALL (src)
1926	1694	1409

For Galactic Plane science, all detections present in 3XMMe are considered potential candidates.

3 Additional information

The enhanced 3XMM catalogue contains some additional information.

- Columns are added for the weighted average fluxes (and their errors) for each unique source. These are provided for each instrument, for energy bands 1-5 and 8 (as defined in http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR5/3XMM-DR5_Catalogue_User_Guide.html#TabBands).

- Columns are added to provide the average (and weighted average) HI (atomic), H_2 (molecular) and total galactic hydrogen column density values and the associated E(B-V) values in the direction of the unique source position. This information is derived from the tool available at http://www.swift.ac.uk/analysis/nhtot/index.php and is based on the analysis of Gamma-ray bursts presented by Willingale et al. (2012). The data added to the enhanced 3XMM catalogue was obtained directly from a command-line version of the tool run at Leicester, kindly provided by Phil Evans. Two further columns of E(B-V) and N_H are provided, based on the reddening analysis of Schlafly & Finkbeiner (2011); these latter

two values supersede those used in previous, internal versions of 3XMMe, which were derived from the work of Schlegel, Finkbeiner & Davis (1998).

- Columns are added containing flags to indicate whether a unique source is considered as a candidate for a given science theme (as described in section 2.2.4). There are separate columns to show when a unique source contains ≥ 1 constituent detections that pass the filtering criteria for that science, when all constituent detections pass the criteria and to show the faction of constituent detections that pass the criteria. These columns are provided for the AGN and Cluster science themes but not for the Galactic theme – all sources are initially considered candidates for Galactic science.

- A column is added to highlight situations where detections lie in fields affected by bright stars or galaxies. Furthers columns indicates where a unique source comprises one or more detections that lie in fields that are affected, where all detections in a unique source lie in such fields and the fraction of detections in the unique source that lie in affected fields.

- It was originally anticipated that the enhanced 3XMM catalogue might be augmented with X-ray spectral parameters derived from XMM-Newton pipeline spectral products as part of the EU-funded PRODEX spectral fit project. However, due to the complexities of providing the data in a meaningful form within the unique source catalogue (since the spectral fit data are provided for each detection), and because the spectral fit database is now public, it was concluded that decisions on how to combine data from the enhanced 3XMM catalogue and the PRODEX spectral fit project should be left to the user. This is possible because the spectral fit catalogue/database contain both the detection ID (DETID) and unique source ID (SRCID) of each detection for which a spectrum was fitted and these identifiers are also contained in the detection (DETID and SRCID) and unique source (SRCID) versions of the enhanced 3XMM catalogue. The spectral fit database (Corral et al. 2015) is accessible at http://xraygroup.astro.noa.gr/Webpage-prodec/.

4 Catalogue contents

Many of the columns in the enhanced 3XMM catalogue are defined in the 3XMM-DR5 catalogue User Guide that is accessible at http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR5/3XMM-DR5_Catalogue_User_Guide.html (see section 4 - "catalogue content and organisation", particularly the Unique Source section). However, the additional columns in the enhanced 3XMM catalogue are described here.

Note that although detections may have been cleaned from the catalogue, thus potentially changing the values of a number of unique source parameters for sources where detections have been removed, the source identifier (SRCID) and the IAU name (IAUNAME) retain the same values as for the 3XMM-DR5 base catalogue since it was decided this facilitates an important back reference to that catalogue. In particular, where detections have been cleaned from unique sources, the SC_RA and SC_DEC values of that source may have changed such that they are no longer consistent with the IAUNAME.

SC_<inst>_<i>_FLUX: The mean flux, for EPIC instrument <inst> (inst = PN, M1 and M2), in band <i> (i=1-5 and 8) of all the detections of the source SRCID, weighted by the errors, in erg/cm^2/s.

SC_EP_EXTENT: The extent, in arcseconds, of the constituent detection of the unique source that has the largest EPIC extent likelihood (reflected in the SC_EP_EXT_ML column). Sources with extent < 6 arcseconds are considered to be insignificantly distinguishable from point sources and have the value reset to 0.0, while an upper limit of 80 arcseconds is imposed. In 3XMM-DR5 this column was labeled SC_EXTENT.

SC_EP_EXTENT_ERR: The error on SC_EP_EXTENT, in arcseconds.

SC_EP_EXT_ML: The all-EPIC log-likelihood value that the source is extended. It is the value for the detection that has the largest EP_EXTENT_ML. In 3XMM-DR5, this column was labelled SC_EXT_ML.

OFFAX: The angle, in arcmins, between the detection (astrometrically corrected coordinates) and the spacecraft boresight of the observation. This parameter is not present in the catalogue of unique sources because its value is not meaningful where constituent detections may have a range of off-axis angles.

A_NH1, W_NH1: simple (A) and weighted (W) averages of the H1 (atomic) hydrogen column density
(see http://www.swift.ac.uk/analysis/nhtot/docs.php).

A_NH2, W_NH2: simple (A) and weighted (W) averages of the H_2 (molecular) hydrogen column density
(see http://www.swift.ac.uk/analysis/nhtot/docs.php).

A_NHTOT, W_NHTOT: simple (A) and weighted (W) averages of the total (atomic +molecular) hydrogen column density
(see http://www.swift.ac.uk/analysis/nhtot/docs.php).

A_EBMV, W_EBMV: simple (A) and weighted (W) averages of the colour excess, E(B-V),
(see http://www.swift.ac.uk/analysis/nhtot/docs.php).

EBMV_DUST: colour excess, E(B-V), based on Schlafly & Finkbeiner (2011) - see http://irsa.ipac.caltech.edu/applications/DUST/. These values replace the original E(B-V) measures provided in 3XMMe v1.2 that were based on Schlegel, Finkbeiner & Davis, (1998).

NH_DUST: Hydrogen column density value based on the relation between N_H and Av, i.e. Nh=1.79e21*Av (Predehl & Schmitt, 1995) where Av is derived from the above-mentioned EBMV_DUST value using Av=Rv x E(B-V) where Rv=3.1 (Schlafly & Finkbeiner, 2011). These values replace the original E(B-V) measures provided in 3XMMe v1.2 that were based on Schlegel, Finkbeiner & Davis, (1998).

SC_AGN_ANY: Set to T(rue) for a unique source if any (≥ 1) constituent detections of that unique source pass the filtering criteria defined for the AGN science theme. Otherwise set to F(alse).

SC_AGN_ALL: Set to T(rue) for a unique source only if *all* constituent detections of that unique source pass the filtering criteria defined for the AGN science theme. Otherwise set to F(alse).

SC_AGN_FRAC: The number of constituent detections in a unique source that pass the filtering criteria for AGN, expressed as a fraction of the total number of constituent detections making up the source. Note that these numbers all refer to the numbers of detections that survived the cleaning stage in the production of the enhanced 3XMM catalogue.

SC_CLUS_ANY: Set to T(rue) for a unique source if any (≥ 1) constituent detections of that unique source pass the filtering criteria defined for the cluster science theme. Otherwise set to F(alse).

SC_CLUS_ALL: Set to T(rue) for a unique source only if *all* constituent detections of that unique source pass the filtering criteria defined for the cluster science theme. Otherwise set to F(alse).

SC_CLUS_FRAC: The number of constituent detections in a unique source that pass the filtering criteria for clusters, expressed as a fraction of the total number of constituent detections making up the source. Note that these numbers all refer to the numbers of detections that survived the cleaning stage in the production of the enhanced 3XMM catalogue.

NEARGS_FLAG: Set to T(rue) for a detection if it lies in an XMM field affected by bright stars or galaxies. Otherwise set to F(alse).

SC_NEARGS_ANY: Set to T(rue) for a unique source if any (≥ 1) constituent detections of that unique source pass lie in XMM fields affected by bright stars or galaxies. Otherwise set to F(alse).

SC_NEARGS_ALL: Set to T(rue) for a unique source only if *all* constituent detections of that unique source lie in XMM fields affected by bright stars or galaxies. Otherwise set to F(alse).

SC_NEARGS_FRAC: The number of constituent detections in a unique source that lie in fields affected by bright stars and galaxies, expressed as a fraction of the total number of constituent detections making up the source.

Appendix A

Twelve sky areas, considered to be part of dedicated observing campaigns ('special' fields) and/or containing large bright X-ray sources,

which are to be avoided for the AGN science theme, are tabulated below - the first 6 rows are the special fields. In the table, the celestial coordinates of the low and high boundaries of each region are shown in hh:mm (RA) and dd:mm (DEC) format respectively. XMM observations are considered to lie inside these regions (and therefore detections in these observations are flagged) if the spacecraft boresight pointing direction of the observation lies within the region boundaries.

Region	**RA (low)**	**RA (high)**	**DEC (low)**	**DEC (high)**
CDF-N	12:35	12:38	+61:51	+62:43
CDF-S	03:30	03:34	-28:11	-27:24
Groth-Westphal	14:15	14:19	+52:02	+52:43
COSMOS	09:57	10:03	+01:20	+03:04
Lockman Hole	10:49	10:56	+57:07	+57:51
Subaru	02:15	02:21	-05:40	-04:18
Virgo	12:23	12:39	+11:27	+13:42
Coma	12:55	13:06	+26:43	+28:24
M31	00:31	01:00	+37:43	+42:37
M33	01:31	01:36	+30:00	+31:22
SMC	00:00	02:00	-76:00	-69:00
LMC	05:00	06:00	-72:00	-65:00

Acknowledgements

The 3XMMe catalogue has been produced via the use of software written and run by Dr. Clive Page.

References

Corral, A., Georgantopoulos, I., Watson, M.G. et al. 2015, A&A, 576, 61
Predehl, P. & Schmitt, J.H.M.M., 1995, A&A, 293, 889
Schlafly, E. F. & Finkbeiner. D. P., 2011, ApJ., 737, 103
Schlegel, D.J., Finkbeiner, D. P. & Davis, M., 1998, ApJ., 500, 525
Willingale, R.Starling, R. L. C.Beardmore, et al., 2012, MNRAS, 431, 394