The XMM-Newton serendipitous survey[*]

VII. The third XMM-Newton serendipitous source catalogue

S. R. Rosen[1], N. A. Webb[2,3], M. G. Watson[1], J. Ballet[4], D. Barret[2,3], V. Braito[1,6], F. J. Carrera[5], M. T. Ceballos[5], M. Coriat[2,3], R. Della Ceca[6], G. Denkinson[1], P. Esquej[1], S. A. Farrell[1], M. Freyberg[7], F. Grisé[8], P. Guillout[8], L. Heil[1], D. Law-Green[1], G. Lamer[10], D. Lin[2,3,12], R. Martino[4], L. Michel[8], C. Motch[8], A. Nebot Gomez-Moran[8], C. G. Page[1], K. Page[1], M. Page[11], M.W. Pakull[8], J. Pye[1], A. Read[1], P. Rodriguez[9], M. Sakano[1], R. Saxton[9], A. Schwope[10], A. E. Scott[1], R. Sturm[7], I. Traulsen[10], V. Yershov[11], and I. Zolotukhin[2,3]

[1] Department of Physics & Astronomy, University of Leicester , Leicester, LE1 7RH, UK
[2] Université de Toulouse; UPS-OMP, IRAP, Toulouse, France
   e-mail: `Natalie.Webb@irap.omp.eu`
[3] CNRS, IRAP, 9 av. Colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France
[4] Laboratoire AIM, CEA-IRFU/CNRS/Université Paris Diderot, Service d'Astrophysique, CEA Saclay, 91191 Gif sur Yvette, France
[5] Instituto de Fisica de Cantabria (CSIC-UC), Avenida de los Castros, 39005 Santander, Spain
[6] INAF-Osservatorio Astronomico di Brera, via Brera 28, I-20121 Milano, Italy
[7] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr., 85748 Garching, Germany
[8] Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France
[9] XMM SOC, ESAC, Apartado 78, 28691 Villanueva de la Cañada, Madrid, Spain
[10] Leibniz-Institut for Astrophysics Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
[11] Mullard Space Science Laboratory, University College London, Holbury St Mary, Dorking, Surrey RH5 6NT, UK
[12] Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, 8 College Road, Durham, NH 03824-2600, U.S.A.

ABSTRACT

Context. Thanks to the large collecting area ($3 \times \sim 1500$ cm^2 at 1.5 keV) and wide field of view (30′ across in full field mode) of the X-ray cameras on board the European Space Agency X-ray observatory *XMM-Newton*, each individual pointing can result in the detection of up to several hundred X-ray sources, most of which are newly discovered objects. As *XMM-Newton* has now been in orbit for more than 15 years, hundreds of thousands of sources have been detected.
Aims. Recently, many improvements in the *XMM-Newton* data reduction algorithms have been made. These include enhanced source characterisation and reduced spurious source detections, refined astrometric precision of sources, greater net sensitivity for source detection and the extraction of spectra and time series for fainter sources, with better signal-to-noise. Thanks to these enhancements, the quality of the catalogue products has been much improved with respect to earlier catalogues. Further, almost 50% more observations are in the public domain compared to 2XMMi-DR3, allowing the *XMM-Newton Survey Science Centre* to produce a much larger and better quality X-ray source catalogue.
Methods. The *XMM-Newton Survey Science Centre* has developed a pipeline to reduce the *XMM-Newton* data automatically. Using the latest version of this pipeline, along with better calibration, a new version of the catalogue has been produced, using *XMM-Newton* X-ray observations made public on or before 2013 December 31. Manual screening of all of the X-ray detections ensures the highest data quality. This catalogue is known as 3XMM.
Results. In the latest release of the 3XMM catalogue, 3XMM-DR5, there are 565962 X-ray detections comprising 396910 unique X-ray sources. For the 133000 brightest sources, spectra and lightcurves are provided. For all detections, the positions on the sky, a measure of the quality of the detection, and an evaluation of the X-ray variability is provided, along with the fluxes and count rates in 7 X-ray energy bands, the total 0.2-12 keV band counts, and four hardness ratios. With an aim to identifying the detections, a cross correlation with 228 catalogues of sources detected in all wavebands is also provided for each X-ray detection.
Conclusions. 3XMM-DR5 is the largest X-ray source catalogue ever produced. Thanks to the large array of data products associated with each detection and each source, it is an excellent resource in which to find new and extreme objects.

Key words. Catalogs – Astronomical data bases – Surveys – X-rays: general

1. Introduction

XMM-Newton (Jansen et al. 2001) is the second cornerstone mission from the *European Space Agency Horizon 2000 programme*. It was launched in December 1999 and thanks to the

~1500 cm^2 of geometric effective area (Turner et al. 2001) for each of the three X-ray telescopes aboard, it has the largest effective area of any X-ray satellite (Longinotti 2014). This fact, coupled with the large field of view (FOV) of 30′, means that a single pointing detects on average 50 to 100 serendipitous X-ray sources (Watson et al. 2009).

For the last 19 years, the *XMM-Newton Survey Science Centre*[1] (SSC), a consortium of 10 European Institutes (Watson et al. 2001) has developed much of the *XMM-Newton Science Analysis Software* (SAS)[2] for reducing and analysing *XMM-Newton* data and created pipelines to perform standardised routine processing of the *XMM-Newton* science data. The XMM SSC has also been responsible for producing catalogues of all of the sources detected with *XMM-Newton*. The catalogues of X-ray sources detected with the three EPIC (Strüder et al. 2001a; Turner et al. 2001) cameras that are placed at the focal point of the three X-ray telescopes, have been designated 1XMM and 2XMM successively (Watson et al. 2009), with incremental versions of these catalogues indicated by successive data releases, denoted -DR in association with the catalogue number. This paper presents the latest version of the XMM catalogue, 3XMM. The original 3XMM catalogue was data release 4 (DR4). The publication of this paper coincides with the release of 3XMM-DR5. This version includes one extra year of data and increases the number of detections by 7%, with respect to 3XMM-DR4. The number of X-ray detections in 3XMM-DR5 is 565962, which translate to 396910 unique X-ray sources. The median flux of these X-ray sources is ~2.4× 10^{-14} erg cm^{-2} s^{-1} (0.2-12.0 keV) and the data taken span 13 years. The catalogue covers 877 square degrees of sky (~2.1% of the sky), if the overlaps in the catalogue are taken into account. 3XMM-DR5 also includes a number of enhancements with respect to the 3XMM-DR4 version which are described in appendix A. The 3XMM-DR5 catalogue is approximately 60% larger than the 2XMMi-DR3 release and five times the current size of the *Chandra* source catalogue (Evans et al. 2010). 3XMM uses significant improvements to the SAS as well as incorporating developments with the calibration. Enhancements include better source characterisation, a lower number of spurious source detections, better astrometric precision, greater net sensitivity and spectra and time series for fainter sources, with better signal-to-noise. These improvements are detailed throughout this paper.

A separate catalogue of ultra-violet and optical sources detected with the *XMM-Newton* Optical Monitor (OM Mason et al. 2001) is also produced in the framework of the *XMM-Newton* SSC and is called the *XMM-Newton Serendipitous Ultraviolet Source Survey* (XMM-SUSS in its original form, with the more recent version named XMM-SUSS2, Page et al. 2012). XMM-SUSS2 contains 5 595 331 detections. They correspond to 4 008 879 sources, of which 692 223 have multiple pointings. This is a complementary catalogue to the 3XMM catalogue, as many of the pointings are similar to those included in 3XMM, even if the FOV of the OM is smaller than the EPIC cameras.

3XMM is also complementary to other recent X-ray catalogues such as the *Chandra* source catalogue mentioned above, and the 1SXPS (Swift-X-ray Telescope (XRT) point source) catalogue (Evans et al. 2014) of 151 524 X-ray point sources detected with the Swift-XRT over eight years of operation. 1SXPS has a sky coverage nearly 2.5 times that of 3XMM, but the effective area of the XRT is less than a tenth of each of the telescopes on board *XMM-Newton* (Longinotti 2014). Other earlier

catalogues include all sky coverage, such as the *ROSAT all-sky survey* (RASS Voges et al. 1999), but the reduced sensitivity of *ROSAT* compared to *XMM-Newton* means that the RASS catalogue contains just 20% the number of sources in 3XMM-DR4. However, the different X-ray source catalogues in conjunction with 3XMM allow searches for long term variability. This is particularly useful in the search for tidal disruption events (e.g. Lin et al. 2011, 2013) and other transient objects such as the best candidate for an intermediate mass black hole, ESO 243-49 HLX-1 (Farrell et al. 2009; Webb et al. 2012). Nonetheless, a wide variety of other sources have also been found thanks to the XMM catalogue, such as many new ultra luminous X-ray sources (Walton et al. 2011), eclipsing polars (Vogel et al. 2008; Ramsay et al. 2009), a peculiar isolated neutron star (Pires et al. 2012), distant luminous X-ray clusters (e.g. Lamer et al. 2008), etc.

Whilst this paper covers the 3XMM catalogues in general, some of the data validation presented was carried out on the 3XMM-DR4 version that was made public on 23rd July 2013. 3XMM-DR4 contains 531261 X-ray detections which relate to 372728 unique X-ray sources, taken from 7427 *XMM-Newton* observations. The paper is structured as follows. Section 2 contains information concerning the observations used in the 3XMM-DR5 catalogue. Section 3 covers the 3XMM data processing and details changes made with respect to previous catalogues (see Watson et al. 2009), such as the exposure selection, the time-dependent boresight implemented, the suppression of minimum ionizing particle (MIP) events, the optimised flare filtering, the improved Point Spread Function (PSF) used for the source detection, new astrometric corrections and the newly derived energy conversion factors (ECFs). We also outline the new source flagging procedure. Section 4 covers the source specific products associated with the catalogue, such as the enhanced extraction methods for spectra and time series and the variability characterisation. Section 5 describes the various screening procedures employed to guarantee the quality of the catalogue and Section 6 outlines the statistical methods used for identifying unique sources in the database. Then, Section 7 describes the procedures used to cross correlate all of the X-ray detections with external catalogues, Section 8 discusses the limitations of the catalogue and Section 9 characterises the enhancement of this catalogue with respect to previous versions, with the potential of the catalogue highlighted by several examples of objects that can be found in 3XMM, in Section 10. Finally, information on how to access the catalogue is given in Section 11, and future catalogue updates are outlined in Section 12, before concluding with a Summary.
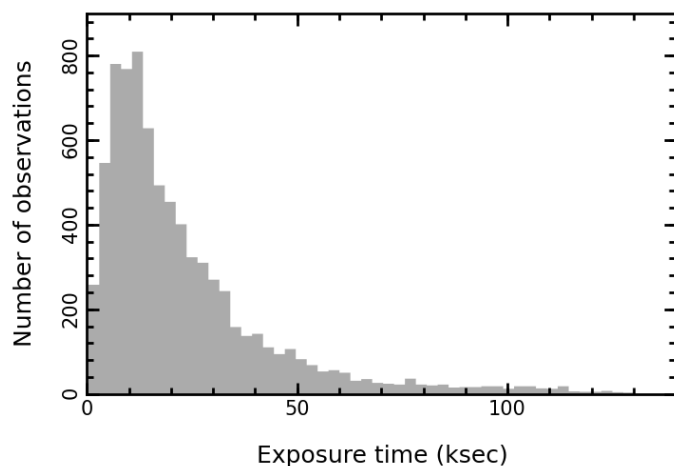
2. Catalogue observations

3XMM-DR5 is comprised of data drawn from 7781 *XMM-Newton* EPIC observations that were publicly available as of 2013 December 31 and that processed normally. The Hammer-Aitoff equal area projection in Galactic coordinates of the 3XMM-DR5 fields can be seen in Fig. 1. The data in 3XMM-DR5 include 440 observations that were publicly available at the time of creating 2XMMi-DR3, but were not included in 2XMMi-DR3 due to the high background or processing problems. All of those observations containing > 1ks clean data (>1 ks of good time interval) were retained for the catalogue. Fig. 2 shows the distribution of total good exposure time (after event filtering) for the observations included in the 3XMM-DR5 catalogue, and using any of the *thick*, *medium* or *thin* filters, but not the *open* filter. The number of the 7781 *XMM-Newton* observations included in the 3XMM-DR5 catalogue for each observing mode and each

---
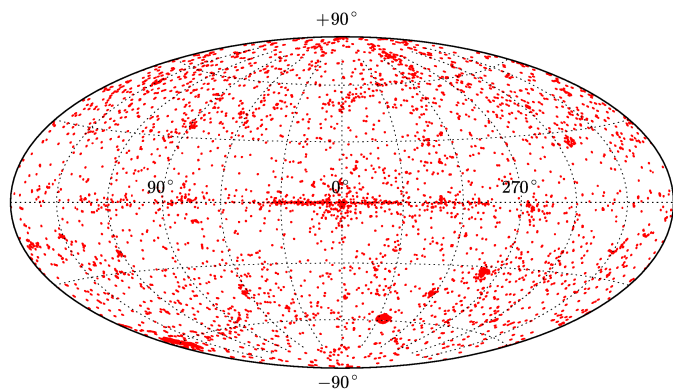
Fig. 1. Hammer-Aitoff equal area projection in Galactic coordinates of the 7781 3XMM-DR5 fields.

Fig. 2. Distribution of total good exposure time (after event filtering) for the observations included in the 3XMM-DR5 catalogue (for each observation the maximum time of all three cameras per observation was used).

filter is given in Table 1. Open filter data were processed but not used in the source detection stage of pipeline processing. The same *XMM-Newton* data modes were used as in 2XMM and are outlined in Watson et al. (2009), their table 1.

The only significant difference was the inclusion of mosaic mode data. Whilst most *XMM-Newton* observations are performed in pointing mode, where the spacecraft is locked on to a fixed position on the sky for the entire observation, since revolution 1812 (2009-Oct-30), a specific mosaic observing mode was introduced in which the satellite pointing direction is stepped across the sky, taking snapshots at points (sub-pointings) on a user-specified grid. Data from dedicated mosaic mode or tracking (mosaic-like) observations are recorded into a single Observation Data File (ODF) for the observation. In previous pipeline processing, the pipeline products from the small number of mosaic-like observations were generally generated, at best, for a single sub-pointing only. This is because the pipeline filters data such that only events taken during an interval where the attitude is stable and centred on the nominal observation pointing direction (within a 3′ tolerance), are accepted. Data from some, or all, of the other sub-pointings were thus typically excluded. During 2012, the *XMM-Newton* Science Operations Centre (SOC) devised a scheme whereby the parent ODF of a mosaic mode observation is split into separate ODFs, one for each mosaic sub-pointing. All relevant data are contained within each sub-

pointing ODF and the nominal pointing direction is computed for the sub-pointing. This approach is applied to both formal mosaic mode observations and those mosaic-like/tracking observations executed before revolution 1812. For a mosaic mode observation, the first 8 digits of its 10-digit observation identifier (OBS_ID) are common for the parent observation and its sub-pointings. However, while the last two digits of the parent observation OBS_ID almost always end in 01, for the sub-pointings they form a monotonic sequence, starting at 31. Mosaic mode sub-pointings are thus immediately recognisable in having OBS_ID values whose last two digits are ≥ 31.

To the pipeline, mosaic mode (and mosaic-like) observation sub-pointings are transparent. No special processing is applied. Each sub-pointing is treated as a distinct observation. Source detection is performed on each sub-pointing separately and no attempt is made to simultaneously fit common sources detected in overlapping regions of multiple sub-pointings. While simultaneous fitting is possible, this aspect had not been sufficiently explored or tested during the preparations for the 3XMM catalogues.

There are 45 observations performed in the dedicated mosaic mode before the bulk processing cut-off date of 2012-Dec-08, of which 37 are included in 3XMM-DR5, see appendix A, point 1. None of these was available for catalogues prior to 3XMM. In total, there are 356 processed mosaic sub-pointings in the 3XMM-DR5 catalogue.

3. Data processing

The data used for the 3XMM catalogues have been reprocessed with the latest version of the SAS and the most up to date calibration available at the time of the processing. The majority of the processing for 3XMM-DR5 was conducted during December 2012/January 2013, with the exception of 20 observations processed during 2013. The SAS used was similar to SAS 12.0.1 but included some upgraded tasks required for the pipeline. The SAS manifest for tasks used in the cat9.0 pipeline and the static set of Current Calibration Files (CCFs) that were used for the bulk reprocessing are provided via a dedicated online webpage[3].

There are 31 observations in 2XMMi-DR3 that did not make it in to 3XMM-DR5, mainly due to software/pipeline errors during processing. Typical examples of the latter problems are due to revised ODFs (e.g. with no useful time-correlation information), more sophisticated SAS software that identified issues hitherto not trapped, or issues with exposure corrections of background flare light curves and pn time-jumps.

The main data processing steps used to produce the 3XMM data products were similar to those outlined in Watson et al. (2009) and described on the SOC webpages[4]. In brief, these steps were the production of calibrated detector events from the ODFs; identification of stable background time intervals; identification of "useful" exposures (taking account of exposure time, instrument mode, etc.); generation of multi-energy-band X-ray images and exposure maps from the calibrated events; source detection and parameterisation; cross-correlation of the source list with a variety of archival catalogues, image databases and other archival resources; creation of binned data products; application of automatic and visual screening procedures to check for any problems in the data products. The data from this process-

---

[3] http://xmmssc-www.star.le.ac.uk/public/pipeline/doc/04_cat9.0_20121220.153
[4] http://xmm.esac.esa.int/sas/current/howtousesas.shtml

Table 1. Characteristics of the 7781 *XMM-Newton* observations included in the 3XMM-DR5 catalogue.

Camera	Modes			Filters			Total
	Full[a]	Window[b]	Other[c]	Thin	Medium	Thick	
pn	5853	495	-	3327	2633	388	6348
MOS1	6045	1306	309	3296	3774	590	7660
MOS2	6100	1341	248	3303	3789	597	7689

[a] Prime Full Window Extended (PFWE) and Prime Full Window (PFW) modes; [b] pn Prime Large Window (PLW) mode and any of the various MOS Prime Partial Window (PPW) modes; [c] other MOS modes (Fast Uncompressed (FU), Refresh Frame Store (RFS)).

ing have been made available through the *XMM-Newton* Science Archive[5] (XSA).

3.1. Exposure selection

The only change applied for identifying exposures to be processed by the pipeline compared to that adopted in pre-cat9.0 processing (Watson et al. (2009) - see their section 4.1), was the exclusion of any exposure taken with the Open filter. This was done because use of the Open filter leads to increased contamination from optical light (optical loading). Eight exposures (from five observations) taken with the Open filter were excluded from the data publicly available for the 3XMM-DR5 catalogue.

3.2. Event list processing

Much of the pipeline processing that converts raw ODF event file data from the EPIC instruments into cleaned event lists has remained unchanged from the pre-cat9.0 pipeline and is described in section 4.2 of Watson et al. (2009). However, we describe 3 alterations to the approach used for 2XMM.

3.2.1. Time-dependent boresight

Analysis by both the *XMM-Newton* SSC and the SOC established the presence of a systematic, cyclic ($\approx$362 day) time-dependent variation in the offset of each EPIC (and OM and RGS) instrument boresight from their nominal pointing positions, for each observation. This seasonal dependence is superposed on a long term trend, the semi-amplitude of the seasonal oscillation being $\approx 1.2''$ in the case of the EPIC instruments (Talavera et al. 2012). These variations of the instrument boresights have been characterised by simple functions in calibration (Talavera et al. 2012; Talavera & Rodríguez-Pascual 2014). The origin of the variation is uncertain but might arise from heating effects in the support structures of the instruments and/or spacecraft star-trackers - no patterns have been identified in the available housekeeping temperature sensor data though these may not sample the relevant parts of the structure.

During pipeline processing of *XMM-Newton* observations for the 3XMM catalogues, corrections for this time-dependent boresight movement are applied to individual event positions in each instrument, via the SAS task *attcalc*, based on the observation epochs of the events.

3.2.2. Suppression of Minimum Ionizing Particle events in EPIC-pn data

High energy particles can produce electron-hole pairs in the silicon substrate of the EPIC-pn detector. While onboard processing and standard pn event processing in the pipeline removes

most of these so-called Minimum Ionizing Particle (MIP) events (Strüder et al. 2001b), residual effects can arise when MIPs arrive during the pre-exposure offset-map analysis and can give rise to features that appear as low-energy noise in the pn detector. Typically, these features are spatially confined to a clump of a few pixels and appear only in band 1. However, in pre-cat9.0 pipeline processing, such features were sometimes detected as sources during source detection and these were not always recognized and flagged during the manual flagging process outlined in section 7.4 of Watson et al. (2009). The SAS task, *epreject* was incorporated into the pipeline processing for 3XMM and in most cases corrects for these MIP events during processing of pn events.

3.2.3. Optimised flare filtering

In previous pipeline processing (pre-cat9.0 pipelines), the recognition of background flares and the creation of Good Time Intervals (GTIs) between them was as described in section 4.3 of Watson et al. (2009), where the background light curves were derived from high energy data and the count rate thresholds for defining the GTIs were based on (different) constant values for each instrument. In the processing for 3XMM, two key changes have been made.

Firstly, rather than adopting fixed count rate thresholds in each instrument, above which data are rejected, an optimisation algorithm has been applied that maximises the signal-to-noise (S/N) for the detection of point sources. Secondly, the light curves of the background data used to establish the count rate threshold for excluding background flares are extracted in an 'in-band' (0.5-7.5 keV) energy range. This was done so that the process described below resulted in maximum sensitivity to the detection of objects in the energy range of scientific interest.

The overall process for creating the background flare GTIs for each exposure within each observation involved the following steps:

1. For each exposure, a high energy light curve (from 7 to 15 keV for pn, > 14 keV for MOS) is created, as previously, and initial background flare GTIs are derived using the optimized approach employed in the SAS task, *bkgoptrate* (see below).
2. Following the identification of bad pixels, event cleaning and event merging from the different CCDs, an in-band image is then created, using the initial GTIs to excise background flares.
3. The SAS task, *eboxdetect* then runs on the in-band image to detect sources with a likelihood > 15 - this is already very conservative as only very bright (likelihood $\gg$ 100), variable sources are able to introduce any significant source variability component into the total count rate of the detector (accumulated from most of the field).
4. An in-band light curve is subsequently generated, excluding events from circular regions of radius 60'' for sources with

---

[5] http://xmm.esac.esa.int/xsa/

count rates $\leq$0.35 counts/s or 100″ for sources with count rates >0.35counts/s, centred on the detected sources.

5. The SAS task, *bkgoptrate*, is then applied to the light curve to find the optimum background rate cut threshold and this is subsequently used to define the final background flare GTIs.

The optimisation algorithm adopted, broadly follows that used for the processing of ROSAT Wide Field Camera data for the ROSAT 2RE catalogue (Pye et al. 1995). The process seeks to determine the background count rate threshold at which the remaining data below the threshold yields a S/N ratio, $S = \frac{C_s}{\sqrt{C_b}}$, for a (constant) source that is a maximum. Here C_s is the number of source counts and C_b is the number of background counts. Since we are interested, here, in finding the background rate cut that yields the maximum S/N and are not concerned about the absolute value of that S/N, then for background light curves with bins of constant width, as created by the pipeline processing, S can be expressed as

$$S = \frac{N}{\sqrt{\sum r_i}} \qquad (1)$$

where N is the number of bins with background count rates below the threshold, r_T, and r_i is the count rate in time bin i: the summation is over the time bins with a count rate $< r_T$. Time bins are of 10s width for pn and 26s for MOS. The process sorts the time bins in order of decreasing count rate. Starting from the highest count rate bin, bins are sequentially removed, computing equation 1 at each step. With the count rate of the bin removed at each step representing a trial background count rate cut threshold, this process yields a curve of S/N vs. background count rate cut threshold. The background cut corresponding to the peak of the S/N curve is thus the optimum cut threshold.

In figure 3 we show four examples of in-band background time series in the top row, accompanied by the respective S/N vs. background-cut-threshold plots in the bottom row. The first panel in each row represents a typical observation (MOS1) with some significant background flaring activity. The optimum cut level of 1.83 cts/s leads to the creation of GTIs that exclude portions of the observation where the background exceeds the cut threshold. The second panels are for a pn observation with a stable, low background level. The optimum cut in the background includes all the data and thus generates a GTI spanning the entire observation. This is also true for the third panels which show a MOS1 case where the background is persistently high (above the level where the whole observation would have been rejected in pre-cat9.0 pipeline processing). The fourth panels are for an example of a variable background which gives rise to a double peaked S/N v background-rate-cut curve. Here, raising the threshold from ~18 cts/s to ~28 cts/s simply involves a steeply rising background rate early in the observation, causing a dip in the S/N verses background-rate-cut curve. However, as the rate cut threshold is increased above 30 cts/s, although the count rate is higher, a lot more exposure time is available, so the S/N curve rises again and the optimum cut includes almost all the data. It should be emphasized that the fixed cut thresholds used for MOS and pn in previous XMM processings can not be directly compared to the optimised ones used here because of the change in energy band being used to construct the background light curve. It is, however, worth noting that the fixed cuts used previously often result in very similar GTIs to those generated by the optimisation process described above. This is because the previous fixed instrument thresholds were based on analyses that sought
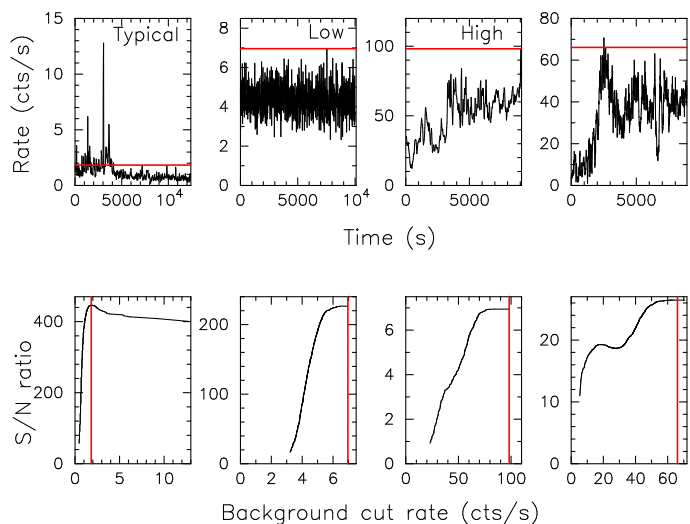
Fig. 3. Flare background light curves (top row) and their corresponding S/N vs. background cut threshold plots (bottom row). The leftmost panels are for a typical observation with notable background flaring. The second pair of vertically aligned panels shows an example where the background has a persistently low level, while the third pair of panels reflects an example where the background is persistently high. The rightmost panels show an example of a variable background which gives rise to a double-peaked S/N vs. background-rate-cut curve. The vertical red lines in the lower panels indicate the optimum background-cut-threshold (i.e. the peak of the curve) derived for the light curves in the top panels. In the upper panels the applied optimum cut-rate is also shown in red as horizontal lines.

to find a representative level for the majority of *XMM-Newton* observations.

We discuss some of the gains of using this optimisation approach in section 9.3 and some known issues in section 8.

3.3. Source detection using the empirical Point Spread Function (PSF) fitting

The bulk reprocessing for 3XMM took advantage of new developments related to the EPIC PSFs. The source detection stage in previous pipelines (Watson et al. (2009) - see their section 4.4.3) made use of the so-called 'default' (or medium accuracy) PSF functions determined by ray tracing of the *XMM-Newton* mirror systems. However, these default PSF functions recognized no azimuthal dependence in the core of the source profile, did not adequately describe the prominent spoke structures seen in source images (arising from the mirror support structures) and were created identically for each EPIC camera.

To address the limitations of the default EPIC PSFs, a set of empirical PSFs were constructed, separately for each instrument, by careful stacking of observed XMM source images over a grid of energy and off-axis angles from the instrument boresights. The cores and spoke patterns of the PSFs were then modelled independently so that implementation within the *XMM-Newton* SAS calibration software then enables PSFs to be reconstructed that take into account the off-axis and azimuthal locations of a source, as well as the energy band. The details of the issues associated with the default PSF and the construction and validation of the empirical PSF are presented in Read et al. (2011).

The use of the empirical PSF has several ramifications in source detection. Firstly, the better representation of structures in the real PSF results in more accurate source parameterisation. Secondly, it helps reduce the number of spurious detec-

tions found in the wings of bright sources. This is because the previous medium accuracy PSFs did not adequately model the core and spoke features, leaving residuals during fitting that were prone to being detected as spurious sources. With the empirical PSFs, fewer such spurious detections are found, especially in the wings of bright objects positioned at larger (> 6′) off-axis angles. Thirdly, as a result of the work on the PSFs, the astrometric accuracy of *XMM-Newton* source positions has been significantly improved (see Read et al. 2011).

3.3.1. Other corrections related to the PSF

During the late stages of testing of the pipeline used for the bulk reprocessing that fed into the 3XMM-DR4 catalogue, an analysis of *XMM-Newton* X-ray source positions relative to the high-accuracy ($\leq 0.1''$) reference positions of SDSS (DR9) quasars identified a small but significant, off-axis-angle-dependent position shift, predominantly along the radial vector from the instrument boresight to the source. The effect, where the real source position is closer to the instrument boresight than that inferred from the fitted PSF centroid, has a negligible displacement on axis and grows to ~0.65″ at off-axis angles of 15′. This radial shift is due to the displacement between the true position of a source and the defined centroid (as determined by a 3-dimensional, circular, Gaussian fit to the model PSF profile) of the empirical PSF, which grows as the PSF becomes increasingly distorted at high off-axis angles. It should be noted that identifying and measuring this effect has only been possible because of the corrections for other effects (see section 3.3 and below) that masked it, and because of the large number of sources available that provide sufficient statistics. In due course a correction for this effect will be applied directly to event positions, on a per-instrument basis, via the *XMM-Newton* calibration system, but for the 3XMM-DR4 catalogue, to avoid delays in its production, a solution was implemented within the *catcorr* SAS task. A correction, computed via a third-order polynomial function, is applied to the initial PSF-fitted coordinates of each source output by *emldetect*, i.e. prior to the field rectification step, based on the off-axis angle of the source as measured from the spacecraft boresight. This correction is embedded in the *RA* and *DEC* columns, which also include any rectification corrections (section 3.4). The correction is computed and applied in the same way for both the 3XMM-DR4 and 3XMM-DR5 catalogues.

A second PSF-related problem that affected 2XMMi-DR3 positions was uncovered during early testing of the empirical PSF (see Read et al. 2011). This arose from a 0.5 pixel error (in both the x and y directions) in the definition of the pixel coordinate system of the medium-accuracy PSF map - as pixels in the PSF map are defined to be 1″ x 1″, the error is equivalent to 0.5″ in each direction. When transferred to the image frame during PSF fitting in *emldetect*, this error in the PSF map coordinate system manifested itself as an offset of up to 0.7″ in the RA/DEC of a source position, varying with azimuthal position within the field. The introduction of the empirical PSF removes this error.

3.4. Astrometric rectification

3.4.1. Frame correction

Celestial coordinates of sources emerging from the PSF fitting step of pipeline processing of a given observation include a generally small systematic error arising from offsets in the spacecraft boresight position from the nominal pointing direction for

the observation. The uncertainty is due to imprecisions in the attitude solution derived from data from the spacecraft's star-trackers and may result in frame shifts that are typically ~1″ (but can be as much as 10″ in a few cases) in the RA and DEC directions and a rotation of the field about the boresight of the order of 0.1 degrees. To correct for (i.e. rectify) these shifts, an attempt is made to cross-correlate sources in the *XMM-Newton* field of view with objects from an astrometric reference catalogue. X-ray sources with counterparts in the reference catalogue are used to derive the frame shifts and rotation that minimise the displacements between them. In all previous pipeline processing (and catalogues derived from them) these frame corrections were estimated using the SAS task, *eposcorr*, which used a single reference catalogue, USNO-B1.0, and the SAS task, *evalcorr*, to determine the success and reliability of the outcome (Watson et al. (2009) - see their section 4.5).

The processing system used to create the data for the 3XMM catalogues makes use of some important improvements to this field rectification procedure, which are embedded in the new SAS tasks, *catcorr* that replaces *eposcorr* and *evalcorr*. Firstly, the new approach incorporates an iterative fitting function (Nelder & Mead 1965) to find the optimum frame-shift corrections: previously the optimum shift was obtained from a grid-search procedure. Secondly, the cross-match between *XMM-Newton* and reference catalogue source positions is carried out using three reference catalogues: (1) USNO-B1.0 (Monet et al. 2003), (2) 2MASS (Skrutskie et al. 2006) and, where sky coverage permits, (3) the Sloan Digital Sky Survey (DR9) (Abazajian et al. 2009). The analysis is conducted using each catalogue separately. When there is an acceptable fit from at least one catalogue, the RA and DEC frame shifts and the rotation derived from the 'best' case are used to correct the source positions. A fit is considered acceptable if there are at least 10 X-ray/counterpart pairs, the maximum offset between a pair (X-ray source, i and counterpart, j) is < 10″ and the goodness of fit statistic

$$L = \sum_{i=1}^{n_x} \sum_{j=1}^{n_o} max(0.0, p_{ij} - q_{ij}) \geq 5 \qquad (2)$$

where $p_{ij} = e^{-\frac{1}{2}(r_{ij}/\sigma_{ij})^2}$ and $q_{ij} = n_o(r_{ij}/r_f)^2$. Here, p_{ij} is the probability of finding the counterpart at a distance > r_{ij} from the X-ray source position given the combined (in quadrature) positional uncertainty, σ_{ij}, while q_{ij} is the probability that the counterpart is a random field object within r_{ij}. An estimate of the local surface density of field objects from the reference catalogue is made by counting the number, n_o, of such objects within a circular region of radius r_f (set to 1′) around each XMM source. n_x is the number of X-ray sources in the XMM field. The L statistic, which represents a heuristic approach to the problem of identifying likely matching counterparts, is computed over ths set of matching pairs and is a measure of the dominance of the closeness of the counterparts over the probability of random matches. The shifts in RA and DEC and the rotation are adjusted within the fitting process to maximise L. Extensive trials found that if $L \geq 5$, the result is generally reliable. Where more than one reference catalogue gives an acceptable solution, the one with the largest L value is adopted.

In the 3XMM catalogues, the corrected coordinates are placed in the RA and DEC columns; the original uncorrected coordinates are reported via the RA_UNC and DEC_UNC columns. A catalogue identifier for the catalogue yielding the 'best' result is provided in the REFCAT column. If the best fit has parameter values (e.g. the number of matches used) that fall

below the specific constraints mentioned above, the original, uncorrected positions are retained (written to both the RA and DEC and RA_UNC and DEC_UNC columns) and the REFCAT identifier takes a negative value. Further details may be found in the documentation for the *catcorr* task. This new rectification algorithm is successful for about 83% of observations, which contain 89% of detections, reflecting a significant improvement compared to the previous approach where ∼ 65% of fields could be corrected. The main gain comes from the use of the 2MASS catalogue which is particularly beneficial in obtaining rectification solutions in the galactic plane - it should be pointed out that similar gains would be obtained with *eposcorr* if used with the expanded set of reference catalogues. It should be noted that the extracted lists of objects from each of the three reference catalogues that lie within the full EPIC field of view for a given observation, are provided to users of *XMM-Newton* data products via the file-type=REFCAT product file, which is used by the task, *catcorr*.

3.4.2. Systematic position errors

As discussed in section 9.5 of Watson et al. (2009), for the 2XMM catalogue (and relevant to subsequent incremental catalogues in the 2XMM series), the angular deviations of SDSS (DR5) quasars (Schneider et al. 2007) from their *XMM-Newton* X-ray counterparts, normalised by the combined position errors, could not be modelled by the expected Rayleigh distribution unless an additional systematic uncertainty (SYSERR parameter in 2XMM) was added to the statistical position error (RADEC_ERR parameter in 2XMM) derived during the PSF fitting process. Watson et al. (2009) showed that this systematic was not consistent with the uncertainty arising from the rectification procedure used for the 2XMM processing and ultimately adopted an empirically-determined systematic error value that produced the best match between the distribution of XMM-quasar offsets and the expected Rayleigh curve.

As part of the upgrade applied to the rectification process for the bulk reprocessing used for the 3XMM catalogues, the uncertainty arising from this step has been computed, in particular, taking into account the error component arising from the rotational offset. Errors (1σ) in each component, i.e., on the RA offset, $\Delta\alpha_c$, on the DEC offset ($\Delta\delta_c$) and on the rotational angle offset ($\Delta\phi_c$), have been combined in quadrature to give an estimate of the total positional uncertainty, Δr, arising from the rectification process as

$$\Delta r = [(\Delta\alpha_c)^2 + (\Delta\delta_c)^2 + (\theta_c.\Delta\phi_c)^2]^{\frac{1}{2}} \quad (3)$$

where θ_c is the radial off-axis angle, measured in the same units as $\Delta\alpha_c$ and $\Delta\delta_c$ and $\Delta\phi_c$ is in radians.

Inclusion of this rectification error (column *SYSERRCC* in the 3XMM catalogues), in quadrature with the statistical error, leads to a generally good agreement between the XMM-quasar offset distribution and the expected Rayleigh distribution compared to the previous approach and indicates that the empirically-derived systematic used in pre-3XMM catalogues is no longer needed. This is discussed further in section 9.2.

3.5. Energy Conversion Factors (ECFs)

A number of improvements in the calibration of the MOS and pn instruments have occurred since the previous, 2XMMi-DR3, catalogue was produced, which lead to slight changes in the Energy

Conversion Factors (ECFs) that are used for converting count rates in the EPIC energy bands to fluxes (see Watson et al. (2009) section 4.6). Of note is the fact that MOS redistribution matrices were provided for 13 epochs at the time of processing for 3XMM and for three areas of the detector that reflect the so-called 'patch', 'wings-of-patch' and 'off-patch' locations (Sembay et al. 2011).

For the 3XMM catalogues a simple approach has been adopted. ECFs were computed following the prescription of Mateos et al. (2009), for energy bands 1 to 5 (0.2-0.5 keV, 0.5-1.0 keV, 1.0-2.0 keV, 2.0-4.5 keV and 4.5-12.0 keV respectively) and band 9 (0.5-4.5 keV), for full-frame mode, for each EPIC camera, for each of the Thin, Medium and Thick filters. A power-law spectral model with a photon index, $\Gamma = 1.7$ and a cold absorbing column density of $N_H = 3 \times 10^{20}$ cm^{-2} was assumed. As such, users are reminded that the ECFs, and hence the fluxes provided in the 3XMM catalogues, may not accurately reflect those for specific sources whose spectra differ appreciably from this power-law model - see section 4.6 of Watson et al. (2009).

For pn, the ECFs are calculated at the on-axis position. The pn response is sufficiently stable that no temporal resolution is needed. For MOS, to retain a direct connection between the ECFs and publicly available response files, the ECFs used are taken at epoch 13 and are for the 'off-patch' location. The latter choice was made because the large majority of detections in an *XMM-Newton* field lie outside the 'patch' and 'wings-of-patch' regions, which only relate to a region of radius ≤ 40″, near the centre of the field. The use of a single epoch (epoch 13) was made to retain simplicity in the processing and because the response of the MOS cameras exhibits a step function change (due to a gain change) between epochs 5 and 6, with different but broadly constant values either side of the step. None of the 13 calibration epochs represent the average response and thus no response file exists to which average ECFs can be directly related. The step-function change in the responses for MOS is most marked in band 1 (0.2-0.5 keV) for the 'patch' location, where the maximum range in ECFs either side of the step amounts to 20%. Outside the 'patch' region, and for all other energy bands, the range of the ECF values with epoch is ≤ 5% and is ≤ 2.5% for the 'off-patch' region. Epoch 13 was chosen, somewhat arbitrarily, as being typical of epochs in the longer post-step time interval.

The ECFs, in units of 10^{11} cts cm^2 erg^{-1}, adopted for the bulk reprocessing of data used for 3XMM, are provided in Table 2, for each camera, energy band and filter. The camera rate, *ca_RATE*, and flux, *ca_FLUX*, are related via *ca_FLUX* = (*ca_RATE*/ECF) (where *ca* is PN, M1 or M2)

3.6. Updated flagging procedures

A significant issue in terms of spurious detections in *XMM-Newton* data arises from detections associated with Out-of-Time (OoT) events. For sources that do not suffer significantly from pile-up, the background map used by emldetect includes a component that models the OoT features. However, for sources where pile up is significant, the OoT modelling is inadequate. This can give rise to spurious sources being detected along OoT features. For the more piled up objects, the numbers of spurious detections along OoT features can become large (tens to hundreds).

Another feature arising from bright sources that affects the MOS instruments is scattered X-rays from the Reflection Grating Arrays (RGA). These manifest themselves as linear features

Table 2. Energy conversion factors (in units of 10^{11} cts cm^2 erg^{-1}) used to convert count rates to fluxes for each instrument, filter and energy band

Camera	Band	Thin	Filters Medium	Thick
pn	1	9.52	8.37	5.11
	2	8.12	7.87	6.05
	3	5.87	5.77	4.99
	4	1.95	1.93	1.83
	5	0.58	0.58	0.57
	9	4.56	4.46	3.76
MOS1	1	1.73	1.53	1.00
	2	1.75	1.70	1.38
	3	2.04	2.01	1.79
	4	0.74	0.73	0.70
	5	0.15	0.15	0.14
	9	1.38	1.36	1.20
MOS2	1	1.73	1.52	0.99
	2	1.76	1.71	1.39
	3	2.04	2.01	1.79
	4	0.74	0.73	0.70
	5	0.15	0.15	0.15
	9	1.39	1.36	1.21

in MOS images passing through the bright object, rather similar in appearance to OoT features. These features are not modelled at all in the background map.

In previous catalogues, spurious detections associated with OoT and RGA features have simply been masked during manual screening. In the cat9.0 pipeline, for the first time, an attempt has been made to identify the presence of OoT and RGA features from piled up sources and to flag detections that are associated with them.

The SAS task, *eootepileupmask*, is used for this purpose. This task uses simple instrument (and mode) -dependent pre-defined thresholds to test pixels in an image for pile-up. Where it detects pixels that exceed the threshold, the column containing that pixel is flagged in a mask map for the instrument. The task attempts to identify and mask columns and rows associated with such pixels in OoT and RGA features.

Once the pile up masks are generated, the SAS task, *dpssflag* is used to set flag 10 of the *PN_FLAG, M1_FLAG, M2_FLAG, EP_FLAG* columns in the catalogues for any detection whose centre lies on any masked column or row.

4. Source-specific product generation

4.1. Optimised spectral and time series extraction

The pipeline processing automatically extracts spectra and time series (source-specific products, SSPs), from suitable exposures, for detections that meet certain brightness criteria.

In pre-cat9.0 pipelines, extractions were attempted for any source which had at least 500 EPIC counts. In such cases, source data were extracted from a circular aperture of fixed radius (28″), centred on the detection position, while background data were accumulated from a co-centred annular region with inner and outer radii of 60″ and 180″, respectively. Other sources that lay within or overlapped the background region were masked during the processing. In most cases this process worked well. However, in some cases, especially when extracting SSPs from sources within the small central window of MOS Small-Window mode observations, the background region could comprise very little usable background, with the bulk of the region lying in the gap

between the central CCD and the peripheral ones. This resulted in very small (or even zero) areas for background rate scaling during background subtraction, often leading to incorrect background subtraction during the analysis of spectra in XSPEC (Arnaud 1996).

For the bulk reprocessing leading to the 3XMM catalogues, two new approaches have been adopted and implemented in the cat9.0 pipeline.

1. The extraction of data for the source takes place from an aperture whose radius is automatically adjusted to maximise the signal-to-noise (S/N) of the source data. This is achieved by a curve-of-growth analysis, performed by the SAS task, *eregionanalyse*. This is especially useful for fainter sources where the relative important of the background level is higher.

2. To address the problem of locating an adequately filled background region for each source, the centre of a circular background aperture of radius, r_b = 168″ (comparable area to the previously used annulus) is stepped around the source along a circle centred on the source position. Up to 40 uniformly spaced azimuthal trials are tested along each circle. A suitable background region is found if, after masking out other contaminating sources that overlap the background circle and allowing for empty regions, a filling factor of at least 70% usable area remains. If none of the background region trials along a given circle yields sufficient residual background area, the background region is moved out to a circle of larger radius from the object and the azimuthal trials are repeated. The smallest trial circle has a radius, r_c, of $r_c = r_b + 60″$ so that the inner edge of the background region is at least 60″ from the source centre - for the case of MOS Small-Window mode, the smallest test circle for a source in the central CCD is set to a radius that already lies on the peripheral CCDs. Other than for the MOS Small-Window cases, a further constraint is that, ideally, the background region should lie on the same instrument CCD as the source. If no solution is found with at least a 70% filling factor, the background trial with the largest filling factor is adopted.

For the vast majority of detections where SSP extraction is attempted, this process obtains a solution in the first radial step and a strong bias to early azimuthal steps, i.e. in most cases an acceptable solution is found very rapidly. For detections in the MOS instruments, about 1.7% lie in the central window in Small-Window mode and have a background region located on the peripheral CCDs. Importantly, in contrast to earlier pipelines, this process always yields a usable background spectrum for objects in the central window of MOS Small-Window mode observations.

This approach to locating the background region was adopted primarily to provide a single algorithm that works for all sources, including those located in the MOS small window, when used. However, a drawback relative to the use of the original annular background region arises where sources are positioned on a notably ramped or other spatially variable background (e.g. in the wings of cluster emission), where the background that is subtracted can vary, depending on which side of the source the background region is located.

In addition, the cat9.0 pipeline permits extraction of SSPs for fainter sources. Extraction is considered for any detection with at least 100 EPIC source counts (EP_8_CTS). Where this condition is met, a spectrum from the source aperture (i.e. source plus background) is extracted. If the number of counts from spectrum

channels not flagged as 'bad' (in the sense adopted by XSPEC) is > 100, a spectrum and time series are extracted for the exposure. The initial filter on EPIC counts is used to limit the processing time as, for dense fields, the above background location process can be slow.

4.2. Attitude GTI filtering

Occasionally, the spacecraft can be settling on to, or begin moving away from, the intended pointing direction within the nominal observing window of a pointed *XMM-Newton* observation, resulting in notable attitude drift at the start or end of an exposure. Image data are extracted from events only within 'Good Time Intervals' (GTIs) when the pointing direction is within $3'$ of the nominal pointing position for the observation. However, in pre-cat9.0 pipelines, spectra and time series have been extracted without applying such attitude GTI filtering. Occasionally, this resulted in a source location being outside or at the edge of the field of view when some events were being collected, leading to incorrect transitions in the time series. In some cases, these transitions gave rise to the erroneous detection of variability in subsequent time series processing. In the cat9.0 pipeline, attitude GTI filtering is applied during the extraction of spectra and time series.

4.3. Variability characterisation

As with pre-cat9.0 pipeline processing, the pipeline processing for the 3XMM catalogues subjects each extracted exposure-level source time series to a test for variability. This test is a simple χ^2 analysis for the null hypothesis of a constant source count rate (Watson et al. (2009) - see their section 5.2). Sources with a probability $\leq 10^{-5}$ of being constant have been flagged as being variable in previous *XMM-Newton* X-ray source catalogues and this same approach is adopted for 3XMM.

In addition, for 3XMM, we have attempted to characterise the scale of the variability through the fractional variability amplitude, F_{var} (provided via the *PN_FVAR, M1_FVAR* and *M2_FVAR* columns and associated *FVARRERR* columns), which is simply the square root of the excess variance, after normalisation by the mean count rate, $\langle R \rangle$, i.e.

$$F_{var} = \sqrt{\frac{(S^2 - \langle \sigma_{err}^2 \rangle)}{\langle R \rangle^2}} \quad (4)$$

(e.g. Edelson et al. (2002); Nandra et al. (1997) and references therein), where S^2 is the observed variance of the time series with N bins, i.e.

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (R_i - \langle R \rangle)^2$$

in which R_i is the count rate in time bin i. For the calculation of the excess variance, $(S^2 - \langle \sigma_{err}^2 \rangle)$, which measures the level of observed variance above that expected from pure data measurement noise, the noise component, $\langle \sigma_{err}^2 \rangle$, is computed as the mean of the squares of the individual statistical errors, σ_i^2, on the count rates of each bin, i, in the time series.

The uncertainty, $\Delta(F_{var})$, on F_{var}, is calculated following equation B2 in appendix B of Vaughan et al. (2003), i.e.

$$\Delta(F_{var}) = \left[\left(\sqrt{\frac{1}{2N}} \frac{\langle \sigma_{err}^2 \rangle}{\langle R \rangle^2 F_{var}} \right)^2 + \left(\sqrt{\frac{\langle \sigma_{err}^2 \rangle}{N}} \frac{1}{\langle R \rangle} \right)^2 \right]^{\frac{1}{2}}$$

This takes account of the statistical errors on the time bins but not scatter intrinsic to the underlying variability process.

5. Screening

As for previous *XMM-Newton* X-ray source catalogues (Watson et al. (2009) - see section 7), every *XMM-Newton* observation in the 3XMM catalogues has been visually inspected with the purpose of identifying problematic areas where source detection or source characterisation are potentially suspect. The manual screening process generates mask files that define the problematic regions. These may be confined regions around individual suspect detections or larger areas enclosing multiple affected detections, up to the full area of the field where serious problems exist. Detections in such regions are subsequently assigned a manual flag (flag 11) in the flag columns (*PN_FLAG, M1_FLAG, M2_FLAG, EP_FLAG*). It should be noted that a detection with flag 11 set to (T)rue does not necessarily indicate that the detection is considered to be spurious.

One significant change to the screening approach adopted for 3XMM relates to the flagging of bright sources and detections within a halo of suspect detections around the bright source. Previously, all detections in the halo region, including the primary detection of the bright source itself (where discernible), had flag 11 set to True (manual flag) but the primary detection of the bright object itself, also had flag 12 set. The meaning of flag 12 there was to signify that the bright object detection was not considered suspect. The use of flag 12 in this 'negative' context, compared to the other flags, was considered to be potentially confusing. For this reason, for the 3XMM catalogues, we have dropped the use of flag 12 and simply ensured that, where the bright object detection is clearly identified, it is un-flagged (i.e. neither flag 11 or 12 are set). Effectively, flag 12 is not used in 3XMM. It should be noted that bright sources that suffer significant pile-up are not flagged in any way in 3XMM (or in previous *XMM-Newton* X-ray source catalogues).

The masked area of each image is an indicator of the quality of the field as a whole. Large masked areas are typically associated with diffuse extended emission, very bright sources whose wings extend across much of the image, or problems such as arcs arising from single-reflected X-rays from bright sources just outside the field of view. The fraction of the field of view that is masked is characterised by the observation class (*OBS_CLASS*) parameter. The distribution of the six observation classes in the 3XMM catalogues has changed with respect to 2XMMi-DR3 (see table 3). The dominant change is in the split of fields assigned observation classes 0 and 1, with more fields that were deemed completely clean in 2XMMi-DR3 having very small areas (generally single detections) being marked as suspect in the 3XMM catalogues. Often these are features that were considered, potentially, to be unrecognised bright pixels, e.g. detections dominated by a single bright pixel in one instrument with no similar feature in the other instruments. It should be emphasised, however, that the manual screening process is unavoidably subjective.

6. Catalogue construction: unique sources

The 3XMM detection catalogues collate all individual detections from the accepted observations. Nevertheless, since some fields

Table 3. 3XMM observation classification (*OBS_CLASS*) (first column), percentage of the field considered problematic (second column) and the percentage of fields that fall within each class for 2XMMi-DR3 and 3XMM-DR5 (third and fourth columns respectively)

OBS CLASS	masked fraction	2XMMi-DR3	3XMM-DR5
0	bad area = 0%	38%	27%
1	0% < bad area < 0.1%	12%	22%
2	0.1% < bad area < 1%	10%	12%
3	1% < bad area < 10%	25%	24%
4	10% < bad area < 100%	10%	11%
5	bad area = 100%	5%	4%

have at least partial overlaps with others and some targets have been observed repeatedly with the target near the centre of the field of view, many X-ray sources on the sky were detected more than once (up to 48 times in the most extreme case). Individual detections have been assigned to unique sources on the sky (i.e. a common unique source identifier, SRCID, has been allocated to detections that are considered to be associated with the same unique source) using the procedure outlined here. The process used in constructing the 3XMM catalogues has changed from that used for the 2XMM series of catalogues (Watson et al. (2009) - see their section 8.1).

The matching process is divided into two stages. The first stage finds, for each detection, all other matching detections within $15''$ of it, from other fields (i.e. excluding detections from within the same observation, which, by definition, are regarded as arising from distinct sources) and computes a Bayesian match probability for each pair as Budavári & Szalay (2008)

$$p_{match} = \left[1 + \frac{1 - p_0}{B \cdot p_0} \right]^{-1} \qquad (5)$$

Here, B, the Bayes factor, is given by

$$B = \frac{2}{(\sigma_1^2 + \sigma_2^2)} \exp - \left[\frac{\psi^2}{2(\sigma_1^2 + \sigma_2^2)} \right] \qquad (6)$$

where σ_1 and σ_2 are the positional error radii of each detection in the pair (in radians) and ψ is the angular separation between them, in radians. $p_0 = N_* / N_1 N_2$ where N_1 and N_2 are the numbers of objects in the sky based on the surface densities in the two fields and N_* is the number of objects common between them. Each of these N values is derived from the numbers of detections in the two fields and are then scaled to the whole sky. The value of N_* is not known, a priori, and in general can be obtained iteratively by running the matching algorithm. However, here we are matching observations of the same field taken with the same telescope at two different epochs so that in most cases, objects will be common. Of course this assumption is affected by the fact that the two observations being considered may involve different exposure times, different instruments, filters and modes used and different boresight positions (with sources within their fields of view being subject to different vignetting factors). To gauge the impact of such effects in determining N_*, trials using an iterative scheme were run, which indicated that taking $N_* = 0.9 min(N_1, N_2)$ provides a good estimate of N_* without the need for iteration. Finally, with all pairs identified and probabilities assigned, pairs with $p_{match} < 0.5$ were discarded.

In the second, clustering stage, a figure-of-merit is computed for each detection, referred to here as the goodness-of-clustering (GoC), which is the number of matches the detection has with

other detections, normalised by the area of its error circle radius (given by POSERR). This GoC measure prioritises detections that lie towards the centre of a group of detections, and are thus likely to be most reliably associated with a given unique source. The list of all detections is then sorted by this GoC value. The algorithm works down the GoC-sorted list and for each detection, the other detections it forms pairs with are sorted by p_{match}. Then, descending this list of pairs, for each one there are four possibilities for assigning the unique source identifiers: i) if both detections have previously been allocated to a unique source and already have the same SRCID, nothing is done, ii) if neither have a SRCID, both are allocated the same, new SRCID, iii) if only one of them has already been assigned a SRCID, the other is allocated the same SRCID, iv) where both detections in the pair have allocated but different SRCIDs, this represents an ambiguous case - for these, the existing SRCIDs are left unchanged but a confusion flag is set for both detections.

This approach is reliable in matching detections into unique sources in the large majority of cases. Nevertheless, there are situations where the process can fail or yield ambiguous results. Examples typically arise in complex regions, such as where spurious sources, associated with diffuse X-ray emission or bright sources, are detected and, by chance, are spatially close to the positions of other detections (real or spurious) in other observations of the same sky region. Often, in such cases, the detections involved will have manual quality flags set (Watson et al. (2009) - see their section 7.5 and also section 5 above).

Other scenarios that can produce similar problems include i) poorly centroided sources, e.g. those suffering from pile-up or optical loading, ii) cases where frame rectification (see 3.4) fails and positional uncertainties are larger than the default frameshift error of $1.5''$ that is adopted for un-rectified fields, iii) sources associated with artifacts such as out-of-time event features arising from bright objects elsewhere in the particular CCD, or residual bright pixels and iv) where multiple detections of sources that show notable proper motion (which is not accounted for in pipeline processing) can end up being grouped into more than one unique source along the proper motion vector. Overall, in 3XMM-DR5, this matching process has associated 239505 detections with 70453 unique sources that comprise more than one detection.

7. External catalogue cross-correlation

The *XMM-Newton* pipeline includes a specific module, the Astronomical Catalogue Data Subsystem (ACDS), running at the *Observatoire de Strasbourg*. This module lists possible multi-wavelength identifications and generates optical finding charts for all EPIC detections. Information on the astrophysical content of the EPIC field of view is also provided by the ACDS.

When possible, finding charts are built using *g*-, *r*- and *i*-band images extracted from the SDSS image server and assembled in false colours. Outside of the SDSS footprint, images are extracted from the Aladin image server. The list of archival astronomical catalogues used during the 3XMM processing includes updated versions of those used for the 2XMM and adds some of the most relevant catalogues published since 2007. A total of 228 catalogues were queried including Simbad and NED. Note that NED entries already included in ACDS catalogues (e.g. SDSS) were discarded.

Among the most important additions are:
i) the Chandra source catalogue version 1.1 (Evans et al. 2010). This release contains point and compact source data extracted from HRC images as well as available ACIS data public at the

Table 4. Cross-matching statistics between 3XMM sources and other catalogues.

Catalogue	Detections	Catalogue	Detections
Chandra src cat.	63,676	Chandra gal.	9,908
SDSS8	129,252	RAVE	219
USNO-B1.0	229,730	IPHAS	38
WISE	454,957	AKARI	5,598
2MASS	36,830	GLIMPSE	35,572
Simbad	204,657	Planck ERCSC	43,136
NED	296,914		

end of 2009. ACDS accesses the Chandra source catalogue using the VO cone search protocol,

ii) the Chandra ACIS survey in 383 nearby galaxies (Liu 2011),

iii) the SDSS Photometric Catalog, Release 8 (Aihara et al. 2011),

iv) the MaxBCG galaxy clusters catalogue from SDSS (Koester et al. 2007),

v) the 2XMMi/SDSS DR7 cross-correlation (Pineau et al. 2011),

vi) the 3rd release of the RAVE catalogue (Siebert et al. 2011),

vii) the IPHAS Hα emission line source catalogue (Witham et al. 2008),

viii) the WISE All-Sky data Release (Cutri & et al. 2012),

ix) the AKARI mid-IR all-sky survey (Ishihara et al. 2010) and version 1.0 of the all-sky survey bright source catalogue (Yamamura et al. 2010),

x) the Spitzer IRAC survey of the galactic center (Ramírez et al. 2008),

xi) the GLIMPSE Source Catalogue (I + II + 3D Churchwell et al. 2009),

xii) the IRAC-24micron optical source catalogue (Surace et al. 2004)

and xiii) the Planck Early Release Compact Source Catalogue (Planck Collaboration et al. 2011).

Table 4 lists, for a selection of archival catalogues, the number of EPIC detections having at least one catalogue entry in the 99.73% (3 Gaussian σ) confidence region.

The cross-matching method used for 3XMM is identical to that applied in the former XMM catalogues. Briefly speaking, ACDS retains all archival catalogue entries located within the 99.73% confidence region around the position of the EPIC detection. The corresponding error ellipse takes into account systematic and statistical uncertainties on the positions of both EPIC and archival catalogue entries. The 3XMM implementation of the ACDS assumes that the error distribution of EPIC positions is represented by the 2-D Gaussian distribution

$$f(\delta_{RA}, \delta_{DEC}) = \frac{1}{2\pi\sigma_{RA}\sigma_{DEC}} e^{-\frac{1}{2}\left(\frac{\delta_{RA}^2}{\sigma_{RA}^2} + \frac{\delta_{DEC}^2}{\sigma_{DEC}^2}\right)}$$

with

$$\sigma_{RA} = \sigma_{DEC} = \sqrt{\text{RADEC\_ERR}^2/2 + \text{SYSERRCC}^2/2}$$

thereby correcting for the overestimated error value used during the 2XMM processing.

ACDS identifications are not part of the 3XMM catalogue fits file but are made available to the community through the XSA and through the XCAT-DB[6], a dedicated interface developed by the SSC in Strasbourg (Motch et al. 2009, Michel et al.

[6] http://xcatdb.unistra.fr

in press). The XCAT-DB also gives access to the entire 3XMM catalogue and to some of the associated pipeline products such as time series and spectra. Quick look facilities and advanced selection and extraction methods are complemented by simple X-ray spectral fitting tools. In the near future, the database will be enriched by the multi-wavelength statistical identifications and associated spectral energy distributions computed within the ARCHES project (Motch & Arches Consortium 2014). Spectral fitting results from the XMMFITCAT database (Corral et al. 2014) are also partially available.

8. Known issues in the catalogue

A number of small but significant issues have been identified that affect the data in the 3XMM catalogues. Two of these affect both the 3XMM-DR4 and 3XMM-DR5 catalogue. The other issues affect only 3XMM-DR4 and are described in appendix A.

1. The optimised flare filtering process (see section 3.2.3) returns a background rate threshold for screening out background flares for each exposure during processing. However, while this process generally works well, when the background level is persistently high throughout the observation, the optimised cut level, while formally valid, can still result in image data with a high background level. In principle, such cases could be identified by testing the cut threshold against a pre-determined benchmark for each instrument. However, this is complicated by the fact that, since the analysis is now measured in-band, apparently high background levels can also arise in fields containing large extended sources. To simplify the process of identifying affected fields, a visual check was performed during manual screening and fields where high background levels were suspected were noted and detections from those fields are flagged in the 3XMM catalogues via the *HIGH_BACKGROUND* column. This screening approach has been somewhat conservative and subjective. A total of 21779 (20625) detections from 568 (552) *XMM-Newton* observations are flagged as such in the 3XMM-DR5: numbers in parentheses are for 3XMM-DR4.

2. A further issue recognized in the 3XMM catalogues is that of detections in the previous 2XMMi-DR3 catalogue that are not present in the 3XMM catalogues. There are 4922 *XMM-Newton* observations that are common between 2XMMi-DR3 and the 3XMM catalogues. However, amongst these observations, there are 53981 detections that appear in 2XMMi-DR3 that are not matched with a detection in the same observation in the 3XMM catalogues within 10″. About 25700 of these were classified as the cleanest (*SUM_FLAG*≤ 1), point-like sources in 2XMMi-DR3 - these are referred to as missing 3XMM detections in what follows. It should be noted that in reverse, there are 63965 detections in the 3XMM catalogues that are in common observations but not matched with a detection in 2XMMi-DR3 within 10″, approximately 33600 of which are classed as being clean and point-like.

   Of the ~ 25700 missing 3XMM detections, up to 8% are found only in the pn band-1 data. Visual inspection of examples and analysis of the pn detector-image data suggests many of these are probably previously unrecognised MIP features, i.e. spurious detections, in 2XMMi-DR3 (see section 3.2.2), though some may well be real, soft sources. A second, difficult-to-quantify percentage (but ≤7%) of the missing 3XMM detections may have detected counterparts

in the 3XMM catalogues but be unmatched within $10''$ due to imperfect astrometry in either the 2XMMi-DR3 and/or 3XMM catalogue. A third component of up to around 3% of the missing 3XMM detections may be detections in 2XMMi-DR3 that are associated with hitherto unrecognised/unflagged detector features - such features become apparent when the missing 3XMM detections are plotted in detector coordinates for each EPIC instrument, after allowing for likely real detections in the same regions that are detected in more than 1 instrument.

Other explanations for the missing 3XMM detections include

 – A small number (<1%) are pairs of visually verified close sources that were separated in 2XMMi-DR3 but found as either a single extended or a single unresolved point source in 3XMM.
 – A small number of cases are likely spurious detections in the wings of bright sources in 2XMMi-DR3 that were erroneously unflagged during the manual screening process for 2XMMi-DR3 and were not detected in 3XMM.

Nevertheless, the above-mentioned explanations account for only a modest fraction (≤20%) of all the clean, point-like missing 3XMM detections. Some 75% of the missing 3XMM detections have EPIC likelihoods in 2XMMi-DR3, L, < 10 (90% have L < 15). It might be thought that the missing 3XMM detections could arise from spurious detections due to random statistical background fluctuations (false positives) in 2XMMi-DR3 - the numbers of such detections, estimated from simulations, was discussed in in section 9.4 of Watson et al. (2009). However, this is not so because although there are notable changes to the pipeline processing between the 2XMMi-DR3 and 3XMM catalogues, the input ODFs and associated event data are generally the same for the common observations, i.e. the data are not independent. As such, the cause of the majority of the missing 3XMM detections remains unclear. However, in comparing 3XMM against 2XMMi-DR3, we need to acknowledge the changes in processing. In particular, the changes to the flare filtering (see section 3.2.3) can result in subtle changes to the background spline model which can impact on the measured detection likelihood of sources. It is likely these changes are at least partly responsible for the numbers of 2XMMi-DR3 detections not found in 3XMM. It should be noted that more detections appear in 3XMM that are not found in common observations in 2XMMi-DR3 - the cause is likely to be similar, with extra objects being found due to the enhancements in sensitivity afforded in the 3XMM catalogues.

9. Catalogue characterisation

9.1. General properties

The 3XMM-DR5 catalogue contains 565962 (531261) detections, associated with 396910 (372728) unique sources on the sky, extracted from 7781 (7427) public *XMM-Newton* observations - numbers in parentheses are for 3XMM-DR4. Amongst these, 70453 (66728) unique sources have multiple detections, the maximum number of repeat detections being 48 (44 for 3XMM-DR4), see fig. 4. 55640 X-ray detections in 3XMM-DR5 are identified as extended objects, i.e. with a core radius parameter, r_{core}, as defined in section 4.4.4 of Watson et al. (2009), > $6''$, with 52493 of these having r_{core} < $80''$. Overall properties in terms of completeness and false detection rates are not
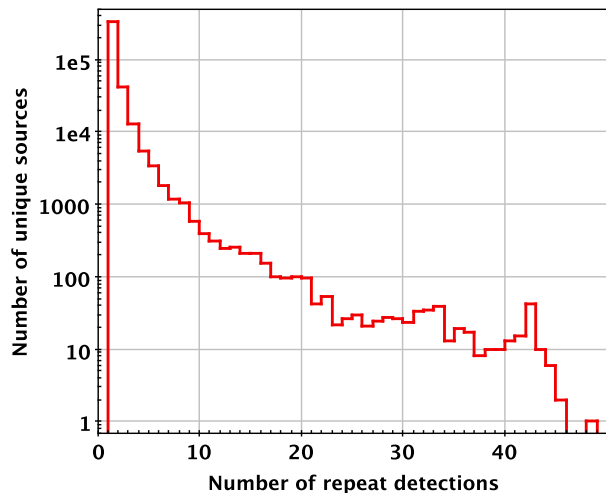

Fig. 4. Numbers of 3XMM-DR5 unique sources comprising given numbers of repeat detections.

expected to differ significantly from those described in Watson et al. (2009).

9.2. Astrometric properties

As outlined in section 3.4, several changes have been made to the processing that affect the astrometry of the 3XMM catalogues relative to previous *XMM-Newton* X-ray source catalogues. To assess the quality of the current astrometry, we have broadly followed the approach outlined in Watson et al. (2009). Detections in the 3XMM-DR5 catalogue were cross-correlated against the Sloan Digital Sky Survey (SDSS) DR12Q quasar catalogue (Paris et al. in prep.), which contains ~297300 objects spectroscopically classified as quasars - positions and errors were taken from the SDSS DR9 catalogue. X-ray detections with an SDSS quasar counterpart within $15''$ were extracted. Point-like 3XMM-DR5 detections were selected with summary flag 0, from successfully *catcorr*-corrected fields, with EPIC detection log-likelihood >8 and at off-axis angles < $13'$. The SDSS quasars were required to have warning flag 0, morphology 0 (point-like) and r' and g' magnitudes both <22.0. This yielded a total of 6614 3XMM-QSO pairs. In the 13 cases where more than one optical quasar match was found within $15''$, the nearest match was retained.

The cross-matching used the *catcorr*-corrected *RA* and *DEC* X-ray detection coordinates. The measured separation, Δr, and the overall 1-dimensional XMM position error, σ_{1D} (= $\sigma_{pos}/\sqrt{2}$), were recorded. Here σ_{pos} is the radial positional error, *POSERR*, in the catalogues, which is the quadrature sum of the XMM positional uncertainties resolved in the RA and DEC directions. As noted by Watson et al. (2009), if the offsets of the X-ray sources and their SDSS quasar counterparts are normalised by the total position error, the distribution of these normalised offsets is expected to follow the Rayleigh distribution,

$$N(x)dx \propto xe^{-x^2/2}dx \qquad (7)$$

where $x = \Delta r/\sigma_{tot}$ - errors on the SDSS quasar positions were included though they are generally ≤ $0.1''$, much smaller than the vast majority of σ_{1D} values in 3XMM-DR5. The
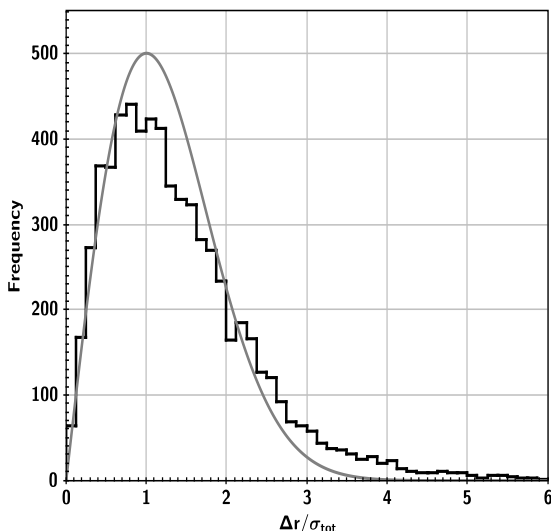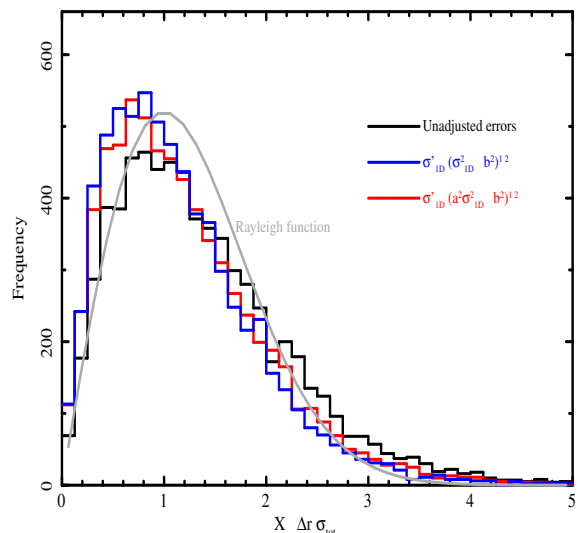
Fig. 5. Distribution of position-error-normalised offsets between 3XMM-DR5 X-ray sources and SDSS quasar counterparts (black histogram). The expected Rayleigh distribution is overlaid (grey). The XMM position errors are as provided in the 3XMM catalogues (i.e. unadjusted, with no scaling or systematic included).

Fig. 6. Similar to figure 5 but comparing results that involve the simplest adjustments to the XMM position errors. For reference, the black histogram is based on using the unadjusted XMM position errors while the expected Rayleigh distribution is overlaid (grey). The blue histogram represents the simplest adjustment to the XMM position errors, involving the addition of a systematic in quadrature, $b(=0.37)$, while the red histogram involves both a scaling of the XMM position error by a factor $a(=1.12)$ and addition of a systematic, $b(=0.27)$, in quadrature. These histograms are based on slightly different filtering compared to figure 5, as explained in the text

SDSS position errors were circularised using $\sigma_{QSO} = [(\sigma_{maj}^2 + \sigma_{min}^2)/2]^{\frac{1}{2}}$ where σ_{maj} and σ_{min} are the errors in the major and minor axis directions of the SDSS position error ellipse. These were then combined in quadrature with the XMM position error to obtain $\sigma_{tot} = (\sigma_{1D}^2 + \sigma_{QSO}^2)^{\frac{1}{2}}$. No systematic error was included for the QSO position error.

In Fig. 5 we show the distribution of x values for the selected XMM-QSO pairs as the black histogram, with the expected Rayleigh distribution overlaid in grey. The normalisation of the Rayleigh curve is determined by the number of pairs used. While there is broad overall agreement between the data and model, it is clear that there is a deficit of sources for $0.8 < x < 2$ and an excess for $x > 2.5$. A total of 739 XMM-QSO pairs lie at $2.5 < x < 6$ while the model predicts 291, the excess of 448 representing 6.8% of the total.

To explore the cause(s) of the deviations from the Rayleigh curve, we began by constructing distributions for many XMM catalogue parameters (e.g. position errors, off-axis angle, count rates, equatorial and galactic location, exposure times, nearest-neighbour distance etc.), comparing the distributions of the data subsets from $x > 3.5$ (the region of the excess where negligible numbers of cases are predicted by the Rayleigh function) and $0 < x < 0.8$ (where the data and model match well). The position error (POSERR) distributions of the two subsets are very similar while the XMM-QSO separations are markedly different, having an average of $0.4''$ for the data from the $0 < x < 0.8$ range compared to $5.5''$ for the $x > 3.5$ subset. There is a weak indication that the points at $x > 3.5$ tend to lie at larger off-axis angles. No other trends could be discerned from the distributions for other parameters. The outlier pairs in the tail may be spurious associations, though as noted by Watson et al. (2009), the false match rate for quasars is expected to yield far fewer spurious associations than the numbers mentioned above for the excess above the Rayleigh curve. Otherwise the results suggest that either the

XMM position errors are being markedly underestimated or the XMM positions are incorrect, for the objects in the excess.

Subsequent analysis investigated whether the discrepancies could be reduced by making phenomenological adjustments to the XMM position errors. In this analysis, the filtering applied to XMM and SDSS sources was similar to that outlined above but only matches within $5''$ were used and no magnitude limits were imposed on the SDSS objects, resulting in 6858 pairs. A two parameter adjustment was considered in which the XMM position errors were scaled by a constant, a, and a systematic error, b, was added in quadrature (i.e. $\sigma'_{1D} = (a^2\sigma_{1D}^2 + b^2)^{\frac{1}{2}}$). One parameter adjustments, where only the systematic was added (i.e. where a is set to 1) were also tested. The error normalised XMM-QSO separations were recomputed as $x' = \Delta r / \sigma'_{tot}$, where σ'_{tot} now combines σ'_{1D} with σ_{QSO} in quadrature. Using this prescription, the data were fit to the Rayleigh function to obtain the best-fit values for a and/or b, using a maximum likelihood approach. The results are shown in figure 6. While these parameterisations of the XMM position errors did improve the fit, particularly bringing the data in the tail closer to the expected Rayleigh curve, the fit remains poor overall, driving the peak of the data to $x \approx 0.7$ (it should peak at 1.0) and introducing a notable excess at $x < 1$. Despite the fact these two forms of adjustment to the XMM position errors yield statistically unacceptable fits to the Rayleigh function, as they do improve agreement in the tail (i.e. for a given XMM-counterpart pair, $x' < x$), they reduce the chance that real matches of 3XMM sources with counterparts from other catalogues (or observations) will be erroneously excluded as candidate counterparts. As such, although the position error column values in the 3XMM catalogues are not adjusted, we provide the values of $a(=1.12)$ and $b(=0.27'')$ for the two parameter fit so that users can apply the above adjustments to the XMM position errors if they wish. The one parameter case best fit yields $b = 0.37$.

Other tests involved (i) imposing a lower bound on the XMM position error ($\sigma'_{1D} = max(\sigma_{1D}, \sigma_{min})$) and (ii) including an off-axis-dependent systematic involving a scalar, c, ($\sigma'^2_{1D} = \sigma^2_{1D} + c^2\Theta^2$) where Θ is the off-axis angle. These latter modifications provide slightly better matches to the Rayleigh curve but still drive the peak of the data to $x \approx 0.7$, again creating an excess at $x < 1$. A further test in which the XMM position error is defined as $\sigma'_{1D} = \sigma_{1D}$ for $x < x_t$ and $\sigma'^2_{1D} = \sigma^2_{1D} + d^2(x - x_t)^2$ for $x \geq x_t$ (where d is a simple scalar and x_t is a threshold value in x) does yield a marked enhancement in the likelihood for the fit but in this case, the data undershoot the Rayleigh curve at $x > 2$ and exceed it at $0.6 < x < 2$.

We conclude that while the more complex adjustments to the XMM position errors can formally improve the match between the error-normalised XMM-QSO separations and the Rayleigh curve, none provides a statistically acceptable match. Moreover, the cases that yield the best improvements in the fit likelihood have no compelling technical rationale.

One possible scenario that might explain the discrepancies is where, in a subset of XMM observations, the *catcorr* task obtains an incorrect solution for the frame rectification. This could manifest itself as erroneous translational (RA/DEC) and/or rotational corrections. A number of XMM fields that contain XMM-QSO pairs from the excess at $x > 3.5$ also contain significant numbers of other XMM-QSO pairs that lie at $x < 3$. If such erroneous rectification corrections were present, we would expect many other pairs in the field to be shifted to larger x values and/or to have a dependence on the off-axis angle (if the rotational correction is wrong). That this is not evident in example fields examined indicates incorrect frame rectification is unlikely to be the underlying cause.
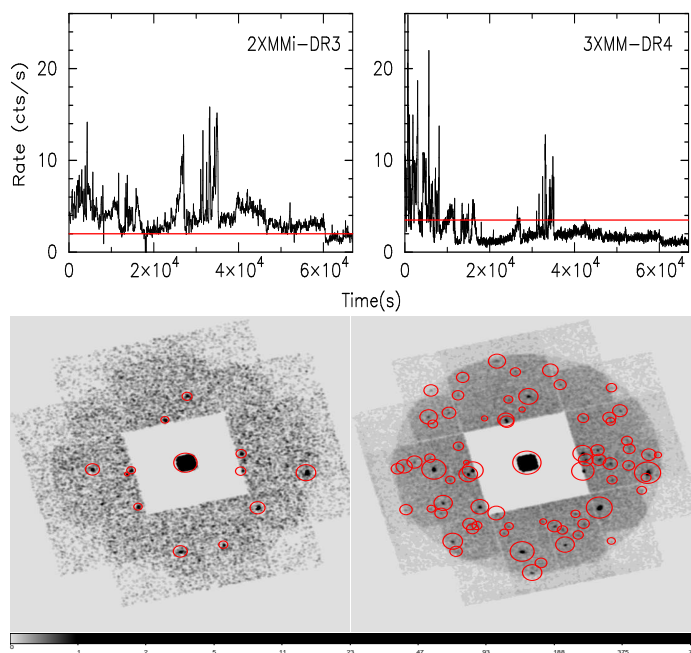
9.3. Background flare filtering

As noted in section 3.2.3, an optimisation algorithm was adopted to determine the count rate threshold for defining the flare GTIs. This process was employed to maximise sensitivity to source detection and can come at the expense of reduced exposure time. Often, the new process results in GTIs that are similar to those derived from the fixed threshold cuts used in pre-cat9.0 pipeline processing. However, in some cases, significant improvements can be obtained in sensitivity.

Of particular interest are cases where the background rises or falls slowly. In such cases, allowing a modest increase in the background count rate can yield a marked increase in exposure time, resulting in a significant improvement in the sensitivity to the detection of faint sources. A good example of this is illustrated in figure 7. As is evident from the light curves, the optimised cut threshold includes significantly more exposure time for a very modest increase in background level, producing a factor 5.5 increase in the harvest of detected sources.

Another aspect of the optimised flare filtering approach is that the increase in exposure time can result in exposures being used that were previously rejected in processing with pre-cat9.0 pipelines.

The pre-cat9.0 and cat9.0 light curves in figure 7 also highlight the fact that the change of energy band used can yield some significant differences in the strengths and even shapes of flare features in the data.

The implementation of the optimised flare filtering approach was done in conjunction with some of the other upgrades, such as the use of the empirical PSF (see section 3.3). As such, we have not directly isolated the impacts on source detection of the

Fig. 7. An example of the improvement offered by the optimised background flare filtering algorithm. Top panels: Left: high-energy MOS1 background flare light curve created by the pre-cat9.0 pipeline, used for the 2XMMi-DR3 catalogue - the red line is the fixed (2 cts/s/arcmin^2) count rate cut threshold applied. Right: in-band (0.5-7.5 keV) light curve used in the cat-9.0 pipeline used for 3XMM-DR4 and 3XMM-DR5 - the red line shows the optimised cut rate threshold derived for the light curve. The lower panels show the resulting, corresponding (smoothed) images, after filtering out the data above the respective rate-cut thresholds. Sources found by the source detection algorithm are indicated by red circles.

optimised flare filtering process alone. Nevertheless, comparison of the numbers of source detections between the set of 4922 observations that are common to the 2XMMi-DR3 and 3XMM-DR5 catalogues, indicates a net increase of 10047 detections in 3XMM-DR5, i.e. a 2.9% increase.

9.4. Extraction of spectral and time series products

As described in section 4.1, spectra and time series of detections are now extracted using optimised extraction apertures that are intended to maximise the overall S/N of the resulting product. To assess this, spectra were re-extracted for all detections and exposures for which spectra were produced during the bulk reprocessing, using a circular aperture of fixed radius (28″) in each case, centred at the same location as the detection position used during the bulk reprocessing. Other than the change of aperture radius, processing was essentially identical to that used in the bulk reprocessing. The S/N, S, of each spectrum was then computed as $S = C_s/C_T^{\frac{1}{2}}$. Here $C_s = C_T - C_b$, where C_T is the total number of counts measured in the spectrum from the source aperture, C_s is the number of counts from the source in the source aperture and C_b is the number of background counts in the source aperture, the latter being estimated from the total counts in the background region, scaled by the ratio of source and background region areas. Counts included in this analysis were drawn from PHA channels with quality < 5 (in XSPEC terms). The S/N was computed in this way for the spectra from the optimised and fixed apertures - the spectral data used for background subtraction were taken from the same background spectrum (from the
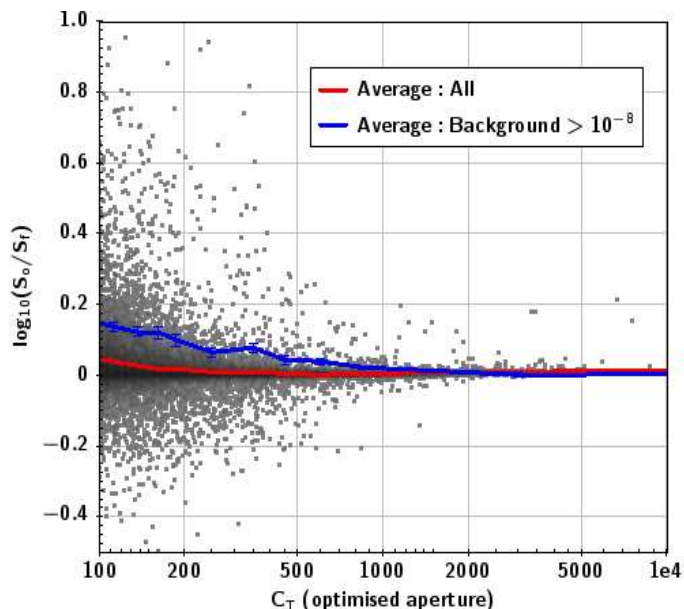
Fig. 8. $log_{10}(S_o/S_f)$ plotted against the log of the total counts, C_T, measured from the optimised aperture. The grey points indicate the data and include only clean (*SUM_FLAG*=0), point-like (*EP_EXTENT* =0) detections. The red line links measurements of the average $log_{10}(S_o/S_f)$, in bins sampling the range in C_T, for cases where $-1 < log_{10}(S_o/S_f) < 1$. The blue line is similar but is for the subset of data where, additionally, the background rate is $> 10^{-8}$ cts s^{-1} (sub-pixel)$^{-2}$ (sub-pixels have side lengths of 0.05 ″). The lower X-axis limit reflects the minimum threshold of 100 total counts in the optimised extraction aperture, imposed for extracting XMM source spectra; the plot is otherwise truncated for clarity.

bulk reprocessing) in each case and the background counts used were drawn only from the same channels as used for the source counts.

In Figure 8 the log of the ratio of the S/N values from the spectra extracted from the optimised (S_o) and fixed (S_f) apertures, i.e. $log_{10}(S_o/S_f)$, is plotted against $log(C_T)$ from the optimised aperture, for MOS1 spectra. Only spectra from the cleanest (*SUM_FLAG*=0), point-like (*EP_EXTENT*=0) detections are included.

It is evident from the positive asymmetry about $log_{10}(S_o/S_f) = 0$, that the optimisation procedure does improve the S/N of the spectra, especially for spectra with lower ($C_T < 500$) numbers of extracted counts, as expected. Overall, 67.5% of the MOS1 spectra with $100 < C_T < 50000$ cts (within $-1 < log_{10}(S_o/S_f) < 1$ which excludes 21 positive outliers) have higher S/N in the optimised aperture than those extracted from the fixed apertures. The red line in figure 8 shows the average of $log_{10}(S_o/S_f)$ of all the data as a function of C_T and indicates that spectra extracted from the optimised apertures with $C_T = 100$ cts have, on average, S/N values 12% higher than those extracted in the fixed apertures. It is anticipated that sources detected in fields with high background levels would benefit from the optimisation procedure. Indeed the blue line in figure 8, which reflects the subset of detections whose background levels are above 10^{-8} cts s^{-1} (sub-pixel)$^{-2}$ (i.e. amongst the highest 15% of background levels), demonstrates this - spectra of such detections extracted from optimised apertures with $C_T = 100$ cts, have average S/N values 39% higher than the spectra from the corresponding fixed apertures.
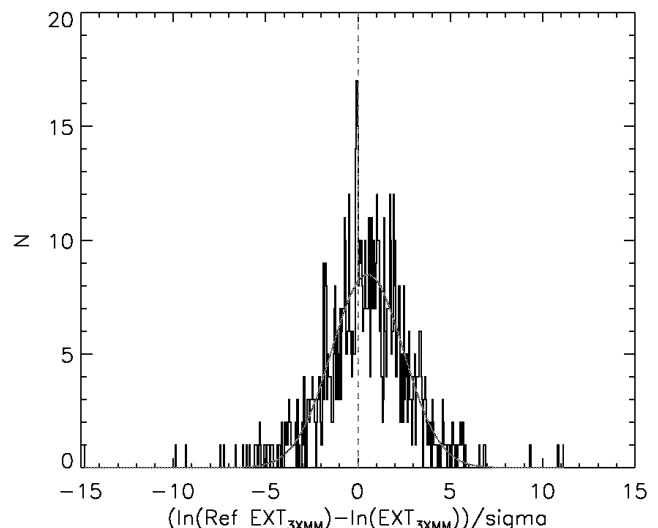

Fig. 10. Histogram of the logarithm of the ratio of extensions between the best observation and the other observations of the same source, normalised by the error. The solid red line is the best Gaussian fit to the histogram. The dashed red line is the expected mean (0).

9.5. Extended sources

The detection and characterisation of extended sources for 3XMM was performed as in 2XMM (Watson et al. 2009). The caveats listed in section 9.9 of that paper still apply to 3XMM. However the better representation of the PSF has helped to improve extended source detection and characterisation. Many extended sources with SC_SUM_FLAG = 4 in 2XMM now have SC_SUM_FLAG = 3 in 3XMM, indicating that the region is still complex but the detection itself is unlikely to be spurious. We have also looked at the distribution of extension likelihood vs. flux as in Fig. 15 of Watson et al. (2009). Fig. 9 shows that 3XMM considers many bright extended sources to be reliable (SC_SUM_FLAG < 2) whereas in 2XMM most of them had higher flag values indicating more significant issues with the data quality.

We have complemented this study by inter-comparing the 3XMM (DR4) results when a source was observed more than once, and with an independent serendipitous search for clusters of galaxies. We restricted the comparison to the best-quality sources with SC_SUM_FLAG = 0. In 3XMM-DR4, 667 sources have been observed several times as extended, each observation being processed independently. We define as the "reference value" the extension (EP_EXTENT) associated to the detection with the highest likelihood value (EP_8_DET_ML column). We investigated the agreement of the extension parameter between the "reference" and the other observations of the same source. We ignored observations when a given source was not detected as extended (mostly because of insufficient exposure) or when the extension was set to 80″ (maximum value allowed in the fit).

In Fig. 10 we show the distribution (in log space) of the ratio between the "reference" extension Ext_{ref} and the current one Ext_{cur}, normalised by the corresponding error equal to $\sqrt{(\sigma_{ref}/Ext_{ref})^2 + (\sigma_{cur}/Ext_{cur})^2}$, where σ_{ref} and σ_{cur} are the extension errors for the "reference" and current observation respectively. We fit the histogram result by a Gaussian function, obtaining a mean value equal to 0.512 (in σ units) with a stan-
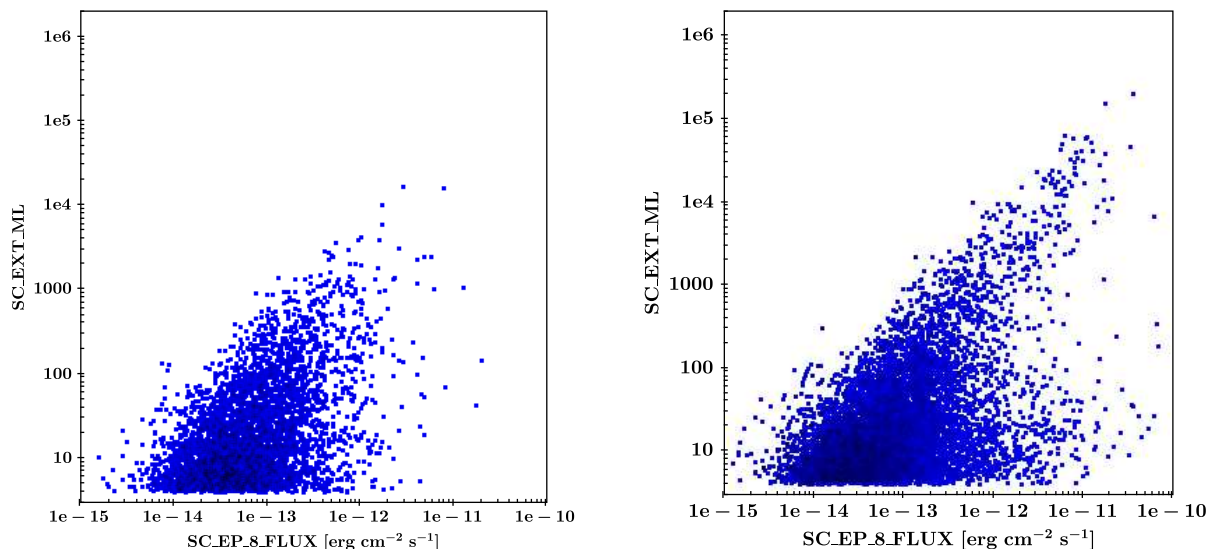
Fig. 9. Distribution of total flux and extension likelihood of all extended sources with SC_SUM_FLAG < 2, in the 2XMM (left) and 3XMM (right) catalogues.
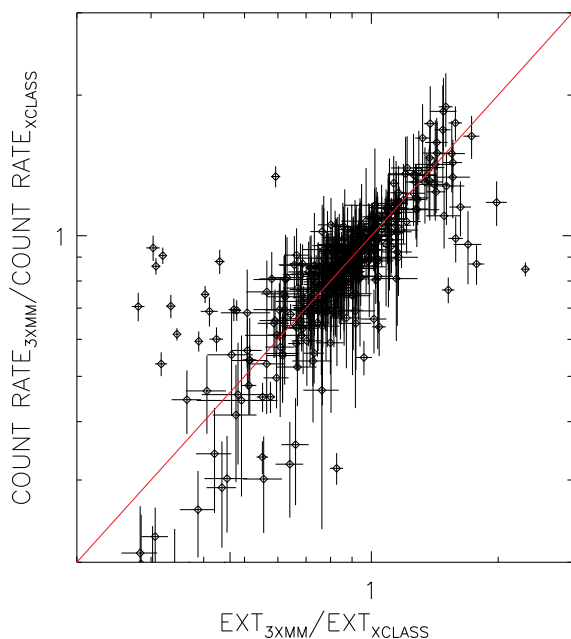
Fig. 11. Comparison of the ratio of extensions and the ratio of count rates obtained by the 3XMM and the XCLASS catalogues. The red solid line is the relation 1:1.

dard deviation equal to 1.943 (we would expect a mean of 0 and a standard deviation of 1 for random fluctuations). We conclude that there exists an additional scatter larger than statistical (of unknown origin) and that the reference observation, which is also the deepest one, estimates a larger extension on average.

The XCLASS catalogue is based on the analysis of archival observations from the XMM-Newton observatory. The XCLASS team processed 2774 high Galactic latitude observations from

the XMM archive (as of 2010 May) and extracted a serendipitous catalogue of some 850 clusters of galaxies based on purely X-ray criteria, following the methodology developed for the XMM Large Scale Survey (Pierre et al. 2007). We used the subsample of 422 galaxy clusters available online at http://xmm-lss.in2p3.fr:8080/l4sdb/ to compare the extension and the count rate obtained for the same sources from the two different procedures (ie. the XCLASS and 3XMM processing). The analytic expression used to represent extended sources in XCLASS was the same as in 3XMM (β-model with β=2/3) so the numbers should be directly comparable. All 422 clusters are in 3XMM-DR4, but 59 (mostly faint or irregular objects) were classified as point sources.

For the 363 extended sources in common, we compared the extent and the count rate in the [0.5-2.0] keV band obtained by 3XMM and XCLASS. We found that, for both quantities, the 3XMM estimates seem to be biased low with respect to the XCLASS values. The best fit regression on source extent resulted in a slope of 0.7 ($Ext_{3XMM} \simeq 0.7 Ext_{XCLASS}$). Excluding clear outliers (difference of extension larger than 20″, typically very faint sources or very bright sources affected by a strong pile-up) the slope increases to 0.85. We conclude that, even excluding these extreme sources, there remains a bias of $\simeq$ 15% between the extensions estimated by 3XMM and XCLASS.

There exists a similar (a little smaller) bias on the count rate. However Fig. 11 shows that there exists a close correlation between both ratios, implying that only one parameter describes the difference in extent and count rates and that, if the source extents were forced to agree, the count rates would agree too. There is no obvious way to know whether the 3XMM or the XCLASS estimate is better but, together with the inter-3XMM comparison, this result indicates that the purely statistical extension error underestimates the real error.

10. Examples

Thanks to the wide range of parameters provided in the catalogue, sources matching specific criteria can be isolated (for ex-

ample variability criteria of X-ray hardness ratios). In this section we show some examples of lightcurves (Fig. 12) and spectra (Fig. 13) extracted from the different EPIC cameras. The plots shown are those associated with the on-line catalogue. Both known and new sources are presented. It is immediately obvious from the two Figures that objects with extremely diverse characteristics are found. Variability on very different timescales is seen in Fig. 12, showing short and long flares, slow rises and steady declines in count rate as well as deep eclipses. From visual examination of the strong variability in Fig. 12c, it was quickly obvious that this new X-ray source was a polar (Webb et al. to be submitted). Fig. 12e shows a strong decline in flux, which, when coupled with the hard spectrum observed for this source, suggests that this might be a previously unknown orphan gamma-ray afterglow.

The spectra shown in Fig. 13 are also very varied and originate from a variety of astrophysical objects, ranging from stars, compact objects, galaxies and clusters of galaxies. An unidentified X-ray source is included in Fig. 13a, which also has a highly variable lightcurve, showing a steady decline in count rate, but with a strong flare superposed. The nature of this source is not obvious and more work will be needed to identify its nature. The sources in the full 3XMM catalogue are of course dominated by unidentified objects, emphasising the large discovery space provided by the catalogue.

11. Catalogue access

The catalogue is provided in several formats. Firstly, a Flexible Image Transport System (FITS) file and a comma-separated values (CSV) file is provided containing all of the detections in the catalogue. For 3XMM-DR5 there are 565962 rows and 323 columns. A separate version of the catalogue (the slim catalogue) is also provided that contains only the unique sources, i.e. 396910 rows, and has 44 columns, essentially those containing information about the unique sources. This catalogue is also provided in FITS and CSV format. Ancillary tables to the catalogue also available from the *XMM-Newton* Survey Science Centre webpages[7] include the table of observations incorporated in the catalogue and the target identification and classification table.

The *XMM-Newton* Survey Science Centre webpages provide access to the 3XMM catalogue, as well as links to the different servers distributing the full range of catalogue products. These include, the *XMM-Newton* XSA, which provides access to all of the 3XMM data products, and the ODF data, the XCat-DB[8] produced and maintained by the *XMM-Newton* SSC, which contains possible EPIC source identification produced by the pipeline by querying 228 archival catalogues. Finding charts are also provided for these possible identifications. Other source properties as well as images, time series, spectra, fit results from the XMM-FITCAT are also provided. Multi-wavelength data taken as a part of the XID (X-ray identification project) run by the SSC over the first fifteen years of the mission are also provided in the XIDresult database[9]. The LEDAS server[10] provides another way to access the 3XMM catalogue and its products, whilst the the upper limit server[11] allows the user to specify a sky position and obtain upper limits on the EPIC fluxes of a point source at

the position if the location has been observed by *XMM-Newton* but no source was detected. The catalogue can also be accessed through HEASARC[12] and VIZIER[13]. The results of the external catalogue cross-correlation carried out for the 3XMM catalogue (section 7) are available as data products within the XSA and LEDAS or through the XCat-DB. The *XMM-Newton* Survey Science Centre webpages also detail how to provide feedback on the catalogue.

Where the 3XMM catalogues are used for research and publications, please acknowledge their use by citing this paper and including the following:

This research has made use of data obtained from the 3XMM XMM-Newton *serendipitous source catalogue compiled by the 10 institutes of the* XMM-Newton *Survey Science Centre selected by ESA.*

12. Future catalogue updates

Incremental releases (data releases) are planned to augment the 3XMM catalogue. An additional year of data will be included with each data release. Data release 6 (DR6) will provide data becoming public during 2014 and should be released by the end of 2015. These catalogues will be accessible as described in Section 11.

13. Summary

This paper presents the third major release of the XMM-Newton serendipitous source catalogue (3XMM), in its original version (3XMM-DR4) and in the first incremental version (3XMM-DR5). The 3XMM catalogues have been constructed by the *XMM-Newton Survey Science Centre* and the 3XMM-DR5 catalogue becomes the largest catalogue of X-ray sources detected using a single X-ray observatory. The characteristics and improvements of this catalogue, with respect to previous versions, are outlined as well as how to cite and access the catalogue. This paper serves as the reference for future incremental versions of the same catalogue (3XMM-DR6, etc), as new *XMM-Newton* data becomes publicly available.

---

[7] http://xmmssc.irap.omp.eu/

[8] http://xcatdb.unistra.fr/3xmm/

[9] http://xcatdb.unistra.fr/xidresult/

[10] http://www.ledas.ac.uk/arnie5/arnie5.php?action=basic&catname=3xmm

[11] http://www.ledas.ac.uk/flix/flix.html

[12] http://heasarc.gsfc.nasa.gov/db-perl/W3Browse/w3table.pl?tablehead=name%3

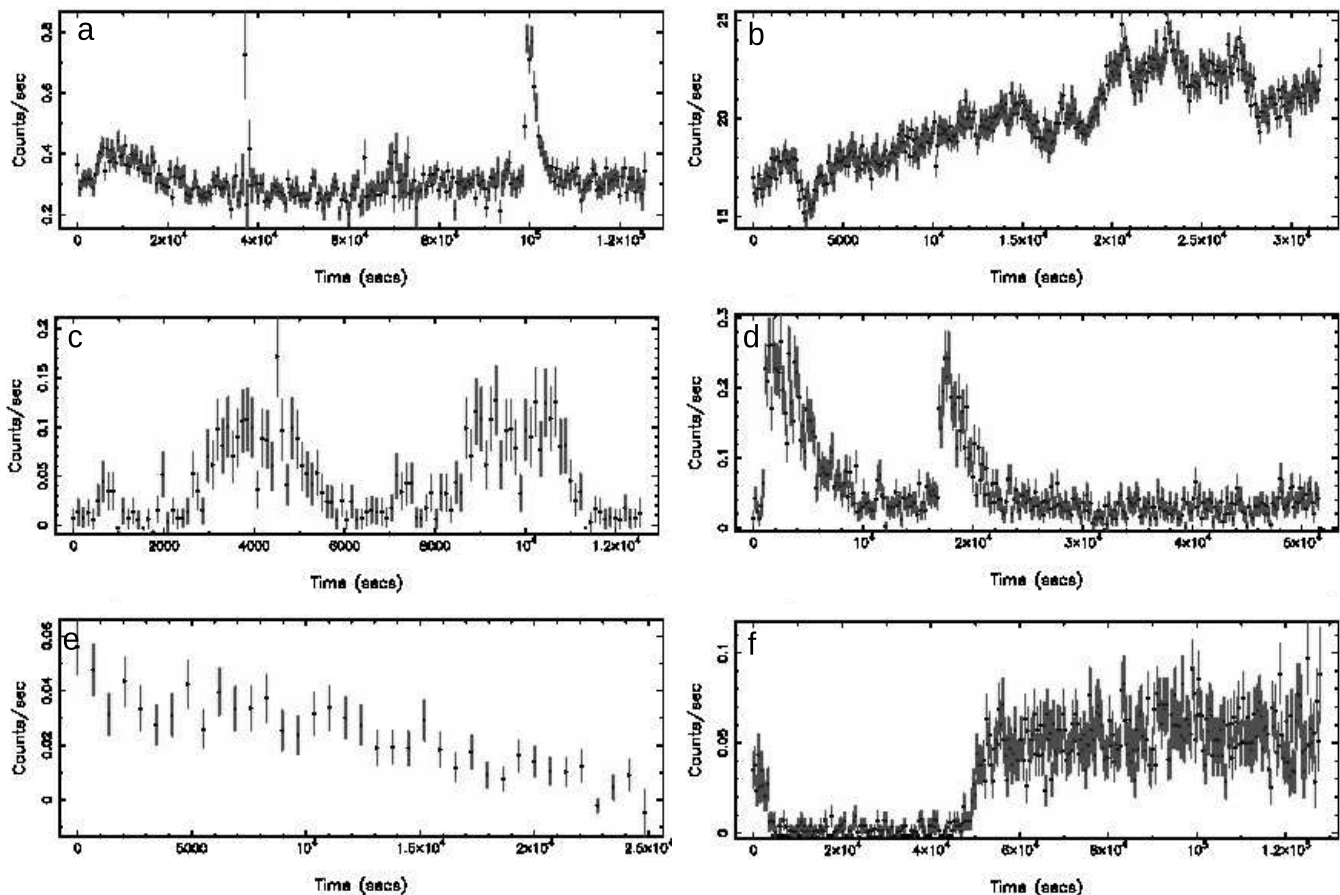[13] http://vizier.u-strasbg.fr/cgi-bin/VizieR

[14] http://heasarc.gsfc.nasa.gov/ftools/

Fig. 12. Example lightcurves taken directly from the 3XMM catalogue. a) 3XMM J111146.1-762010 = CHX 18N – T Tau-type star showing a short flare b) 3XMM J000619.5+201210, A Seyfert 1, Markarian 335. c) 3XMM J184916.1+652943, a new 1.6 hour polar (Webb et al. to be submitted). d) A 2MASS star (2MASS J00025638-3004447) showing two large flares. e) 3XMM J002159.4+614254, a new X-ray object showing a strong decline in flux. Possibly a gamma-ray burst afterglow. f) 3XMM J013334.0+303211, a high mass X-ray binary in M 33, M33 X-7, showing a 12.5 hour eclipse - the first eclipsing stellar-mass black hole binary discovered (Pietsch et al. 2006)

Appendix A: Known issues affecting 3XMM-DR4 only

– After the creation of the 3XMM-DR4 catalogue, it was discovered that the raw event files from the ODFs of a number of mosaic mode sub-pointing observations contained corrupted data whereby some of the events in a given sub-pointing ODF were actually from another sub-pointing. Since the raw event positions are specified in detector coordinates and are subsequently mapped to their sky locations during pipeline processing by reference to the observation boresight position, which is specified for the given sub-pointing, the celestial positions of these events are wrong and therefore results in some detections having incorrect celestial coordinates. The problem arose in the algorithm used to split the raw parent ODF into sub-pointing ODFs. In some cases all instruments were affected while in others, only one or both of the MOS instruments was affected. Of the 419 mosaic-mode sub-pointing observations included in 3XMM-DR4, 82 are affected to some extent, involving 4918 detections. The affected observations are listed in the *watchout* section

of the XMMSSC 3XMM-DR4 catalogue web pages[15]. For 3XMM-DR5, none of the affected mosaic sub-pointing observations is included in the catalogue.

– The vignetting values provided in the 3XMM-DR4 catalogue (for each instrument, for bands 1 to 5) were found to have been computed for an energy of 0 keV rather than the energy relevant to the band. Thus the values for each band of a given instrument are identical. This error does not affect the count rates or fluxes as the vignetting correction applied to them is computed separately and has been verified as correct. It is only the tabulated values in the vignetting columns of the catalogue that are incorrect in 3XMM-DR4 and they are correct in 3XMM-DR5.

– A significant issue identified after the public release of the 3XMM-DR4 catalogue relates to the error values on various quantities. It was established that the error quantities (i.e. columns containing an _ERR at the end) for the XID band (band 9) count rates and fluxes of a significant number (~42200) of detections (~10% of the catalogue) were substantially wrong (generally being overestimated by factors up to ~ 100 but in a few cases, up to 1000). A more

---

[15] http://xmmssc-www.star.le.ac.uk/Catalogue/xcat_public_3XMM-DR4.html
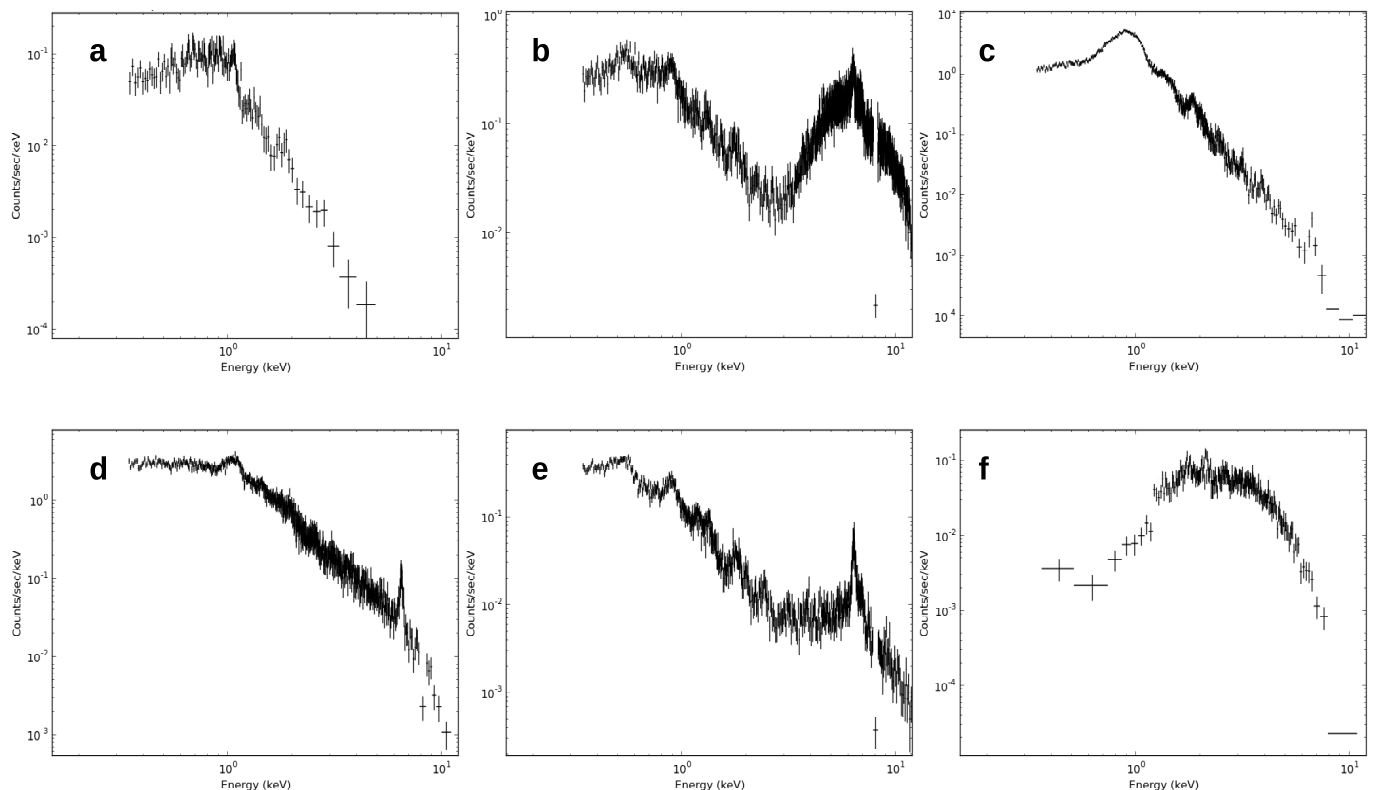
Fig. 13. Example spectra taken directly from the 3XMM catalogue, showing the diversity of the sources in the 3XMM catalogue. Energy (keV) is given on the abscissa and count s^{-1} keV^{-1} on the ordinate. a) 3XMM J052532.5+062533, an X-ray source of unknown nature, as the majority of the sources are in 3XMM b) 3XMM J123536.6-395433 a Seyfert 2 galaxy (NGC 4507) c) 3XMM J125141.9+273226, a rotationally variable star, 31 Com d) 3XMM J162838.2+393303, a cluster of galaxies e) 3XMM J011127.5-380500, the pn spectrum of NGC 424, a Seyfert 2 galaxy f) 3XMM J185246.6+003317, a new transient magnetar discovered by Zhou et al. (2014)

detailed investigation found that while all error columns are potentially affected (and therefore also any derived parameters involving error-weighted quantities, such as some of the unique source quantities), the frequency and magnitude of the problem is much worse for the XID band data than any other parameter. It has been established that for other key quantities, such as the statistical positional uncertainty (*RADEC_ERR*) and the instrument count rates and fluxes in other (non-XID) bands, only about 1.3% of detections are affected and, generally, the scale of the problem is very small. For the positional uncertainty, 1.4% of detections have incorrect *RADEC_ERR* values and only 0.26% of detections have position errors that differ from their correct values by more than 0.05″ while for only 89 detections does it differ by more than 0.5″ (of which, 58 are detected as extended sources and 81 have a non-zero quality flag). Furthermore, for 81% of those detections where the position error is wrong by more than ±0.05″, the correct position error is smaller than that quoted in the 3XMM-DR4 catalogue. The most extreme deviations of the *RADEC_ERR* values from their correct values are 32″ larger and 2.3″ smaller. For the PN band 2 flux errors, only ~1.1% of detections have values that deviate from their correct values by more than 10^{-5}, when expressed as a fraction of the correct value. For the errors on the XID band photometric quantities (rates, fluxes, counts) the correct error is generally smaller than that given in 3XMM-DR4.

Thus, while there is a significant problem with the error quantities on the XID band photometric data in 3XMM-DR4, the problem is much less severe for other quantities. It is emphasized that the correct error quantities are present in 3XMM-DR5.

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
Arnaud, K. A. 1996, in Astronomical Society of the Pacific Conference Series, Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes, 17
Blackburn, J. K. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, 367
Budavári, T. & Szalay, A. S. 2008, ApJ, 679, 301
Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, PASP, 121, 213
Corral, A., Georgantopoulos, I., Watson, M. G., et al. 2014, ArXiv e-prints
Cutri, R. M. & et al. 2012, VizieR Online Data Catalog, 2311, 0
Edelson, R., Turner, T. J., Pounds, K., et al. 2002, ApJ, 568, 610
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37
Evans, P. A., Osborne, J. P., Beardmore, A. P., et al. 2014, ApJS, 210, 8
Farrell, S. A., Webb, N. A., Barret, D., Godet, O., & Rodrigues, J. M. 2009, Nature, 460, 73
Ishihara, D., Onaka, T., Kataza, H., et al. 2010, A&A, 514, A1
Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1
Koester, B. P., McKay, T. A., Annis, J., et al. 2007, ApJ, 660, 239
Lamer, G., Hoeft, M., Kohnert, J., Schwope, A., & Storm, J. 2008, A&A, 487, L33
Lin, D., Carrasco, E. R., Grupe, D., et al. 2011, ApJ, 738, 52
Lin, D., Irwin, J. A., Godet, O., Webb, N. A., & Barret, D. 2013, ApJ, 776, L10
Liu, J. 2011, ApJS, 192, 10
Longinotti, A. L. 2014, XMM-Newton Users Handbook, Tech. Rep. 2.12, ESA: XMM-Newton SOC
Mason, K. O., Breeveld, A., Much, R., et al. 2001, A&A, 365, L36

Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, AJ, 125, 984

Motch, C. & Arches Consortium. 2014, in The X-ray Universe 2014, 288

Motch, C., Michel, L., & Pineau, F.-X. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 466

Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, ApJ, 476, 70

Nelder, J. A. & Mead, R. 1965, The Computer Journal, 7, 308

Page, M. J., Brindle, C., Talavera, A., et al. 2012, MNRAS, 426, 903

Pierre, M., Chiappetti, L., Pacaud, F., et al. 2007, MNRAS, 382, 279

Pietsch, W., Haberl, F., Sasaki, M., et al. 2006, ApJ, 646, 420

Pineau, F.-X., Motch, C., Carrera, F., et al. 2011, A&A, 527, A126

Pires, A. M., Motch, C., Turolla, R., et al. 2012, A&A, 544, A17

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A7

Pye, J. P., McGale, P. A., Allan, D. J., et al. 1995, MNRAS, 274, 1165

Ramírez, S. V., Arendt, R. G., Sellgren, K., et al. 2008, ApJS, 175, 147

Ramsay, G., Rosen, S., Hakala, P., & Barclay, T. 2009, MNRAS, 395, 416

Read, A. M., Rosen, S. R., Saxton, R. D., & Ramirez, J. 2011, A&A, 534, A34

Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, AJ, 134, 102

Sembay, S., Saxton, R. D., & Guainazzi, M. 2011, EPIC MOS response, Tech. Rep. XMM-CCF-REL-272, ESA

Siebert, A., Williams, M. E. K., Siviero, A., et al. 2011, AJ, 141, 187

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Strüder, L., Briel, U., Dennerl, K., et al. 2001a, A&A, 365, L18

Strüder, L., Briel, U., Dennerl, K., et al. 2001b, A&A, 365, L18

Surace, J. A., Shupe, D. L., Fang, F., et al. 2004, VizieR Online Data Catalog, 2255, 0

Talavera, A. & Rodríguez-Pascual, P. 2014, Astrometry: time-dependent boresight update, Tech. Rep. XMM-CCF-REL-315, ESA

Talavera, A., Rodríguez-Pascual, P., & Guainazzi, M. 2012, Astrometry: time-dependent boresight, Tech. Rep. XMM-CCF-REL-286, ESA

Turner, M. J. L., Abbey, A., Arnaud, M., et al. 2001, A&A, 365, L27

Vaughan, S., Edelson, R., Warwick, R. S., & Uttley, P. 2003, MNRAS, 345, 1271

Vogel, J., Byckling, K., Schwope, A., et al. 2008, A&A, 485, 787

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

Walton, D. J., Roberts, T. P., Mateos, S., & Heard, V. 2011, MNRAS, 416, 1844

Watson, M. G., Auguères, J.-L., Ballet, J., et al. 2001, A&A, 365, L51

Watson, M. G., Schröder, A. C., Fyfe, D., et al. 2009, A&A, 493, 339

Webb, N., Cseh, D., Lenc, E., et al. 2012, Science, 337, 554

Witham, A. R., Knigge, C., Drew, J. E., et al. 2008, MNRAS, 384, 1277

Yamamura, I., Makiuti, S., Ikeda, N., et al. 2010, VizieR Online Data Catalog, 2298, 0

Zhou, P., Chen, Y., Li, X.-D., et al. 2014, ApJ, 781, L16