Draft version November 26, 2019
Typeset using LaTeX twocolumn style in AASTeX63

2SXPS: An improved and expanded Swift X-ray telescope point source catalog

P. A. Evans,[1] K.L. Page,[1] J.P. Osborne,[1] A.P. Beardmore,[1] R. Willingale,[1] D.N. Burrows,[2] J.A. Kennea,[2] M. Perri,[3,4] M. Capalbi,[5] G. Tagliaferri,[6] and S.B. Cenko[7]

[1] University of Leicester, X-ray and Observational Astronomy Group, School of Physics and Astronomy, University Road, Leicester, LE1 7RH, UK
[2] Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802, USA
[3] ASI Space Science Data Center, Via del Politecnico, I-00133 Rome, Italy
[4] INAF-Osservatorio Astronomico di Roma, Via Frascati 33, I-000040 Monteporzio Catone, Italy
[5] INAF – IASF Palermo, via Ugo La Malfa 153, I-90146, Palermo, Italy
[6] INAF-Osservatorio Astronomico di Brera, via E. Bianchi 46, 23807 Merate (LC), Italy
[7] NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA

ABSTRACT

We present the 2SXPS (Swift-XRT Point Source) catalog, containing 206,335 point sources detected by the *Swift* X-ray Telescope (XRT) in the 0.3–10 keV energy range. This catalog represents a significant improvement over 1SXPS, with double the sky coverage (now 3,790 deg^2), and several significant developments in source detection and classification. In particular, we present for the first time techniques to model the effect of stray light – significantly reducing the number of spurious sources detected. These techniques will be very important for future, large effective area X-ray mission such as the forthcoming *Athena* X-ray observatory. We also present a new model of the XRT point spread function, and a method for correctly localising and characterising piled up sources. We provide light curves – in four energy bands, two hardness ratios and two binning timescales – for every source, and from these deduce that over 80,000 of the sources in 2SXPS are variable in at least one band or hardness ratio. The catalog data can be queried or downloaded via a bespoke web interface at https://www.swift.ac.uk/2SXPS, via HEASARC, or in Vizier (IX/58).

Keywords: Catalogs – Surveys – X-rays: observations – Methods: data analysis

1. INTRODUCTION

Serendipitous source catalogs have, for many years, been a standard product of X-ray observatories giving great insights into the nature and range of X-ray emitting objects in the Universe. Typically they can be divided into two categories: large area but relatively shallow (such as the *ROSAT* All-sky Survey, RASS: Voges et al. 1999; Boller et al. 2016), or small area but deep (e.g. the *XMM-Newton* catalogs, Watson et al. 2009; Rosen et al. 2016; Traulsen et al. 2019, and the *Chandra* catalogs, Evans et al. 2010). The output of the X-ray telescope (XRT; Burrows et al. 2005) on the *Swift* satellite (Gehrels et al. 2004) lies between these two extremes, and three point-source catalogs have been produced from XRT data. SwiftFT (Puccetti et al. 2011) focussed on the deepest (> 10 ks) datasets, and 1SWXRT (D'Elia et al. 2013) analysed the individual observations; 1SXPS (Evans et al. 2014, hereafter 'paper I') contained analysis of both individual observations, and the combination of multiple, overlapping datasets. 1SXPS covered 1905 square degrees (nearly double that of the more recent 3XMM-DR8 catalog), with a median 0.3–10 keV source flux of 3.0×10^{-14} erg cm^{-2}s^{-1}, compared to 2.2×10^{-14} erg cm^{-2}s^{-1} (0.2–12 keV) in 3XMM-DR8. Although XRT has a lower effective area (100 cm^2 at 1.5 keV) and smaller field of view (radius 12.3′) than *XMM*, it also has a much lower background due to the orbital environment, which recovers much of the comparitive sensitivity. Additionally, *Swift* observes a much larger number of targets than is typical for a satellite, typically carrying out from tens to hundreds of distinct pointings every day.

As well as providing a survey of moderate width and depth, the *Swift*-XRT data provide insight into the variability of the X-ray sky, since 95% of its observations are of areas of the sky which it has observed multiple times.

pae9@leicester.ac.uk

Such information is critical in the current era of time-domain astronomy, particularly multi-messenger astronomy, to aid in the identification of X-ray counterparts to time-domain signals found at other wavelengths or using non-photon triggers. For example, the localisations of astrophysical neutrinos (IceCube Collaboration 2013) or gravitational waves (Singer et al. 2014) are poor and many X-ray sources are found in follow-up observations (Evans et al. 2015, 2016a, 2017). In order to correctly identify the true counterpart from the unrelated sources, an understanding of the temporal properties of the serendipitous X-ray sky is crucial.

In this paper we present an updated *Swift*-XRT point source catalog: 2SXPS. This catalog contains 50% more temporal coverage than 1SXPS, but contains 80% more exposure (Table 1), due to a change in which observations were selected for inclusion (Section 2).

As well as updating the data in the catalog, we have updated our source detection system, focusing particularly on reducing the number of spurious detections due to diffuse emission or stray light, as is discussed in some detail in Section 3.2.

After we had begun processing the data for 2SXPS, Traulsen et al. (2019) produced a catalog based on stacking multiple co-located *XMM* observations. As part of this work they demonstrated the use of an adaptive smoothing technique combined with source masking as a means of estimating the background, which they deemed more reliable in the presence of diffuse or structured emission than the approach previously followed by the 2/3XMM and upon which the 1SXPS background modelling was based. Because *Swift* is in a low-Earth orbit, observations are comprised of one or more 'snapshots' (i.e. continuous exposures in a single orbit) of no more than 2.7 ks in duration. Such snapshots are not perfectly aligned, meaning the background must be modelled for each snapshot individually. The shortness of the snapshots, combined with the smaller effective area of XRT compared to *XMM*-EPIC, and the different orbital environment of the two satellites results in a much lower background in the individual XRT snapshots than in *XMM* observations, and we found in paper I that directly applying the *XMM* background-mapping approach was not appropriate for *Swift*. Since the 2SXPS software and simulation work was in an advanced state when Traulsen et al. (2019) was published, we elected not to investigate applying their technique to XRT data for this work.

2. DATA SELECTION, FILTERING AND STACKED IMAGE CREATION.

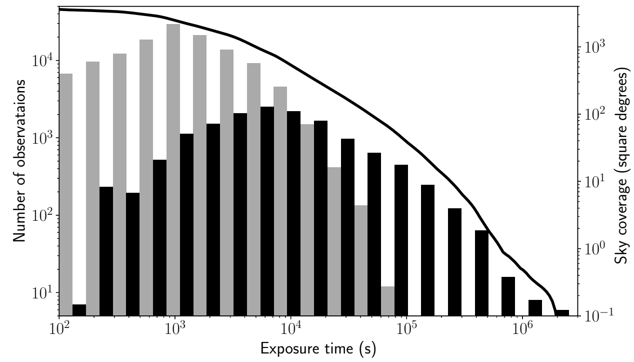

Figure 1. The sky coverage and exposure details of the 2SXPS catalog. The solid line shows the sky coverage (corrected for overlaps) as a function of exposure time. The histogram shows the distribution of exposure time per dataset, with the individual observations shown in light gray and the stacked images in black.

We selected all[1] observations taken between 2005 January 01[2] and 2018 August 01 with at least 100 s of cleaned Photon Counting (PC) mode exposure. For the anlaysis in this catlaog we used XRTDAS[3] v3.4.0 within HEASOFT v6.22, and the most recent XRT calibration as of 2018 August 01. All event lists were reprocessed using XRTPIPELINE to give a self-consistent and up-to-date dataset.

The observations were filtered to remove times where the data were contaminated by scattered light from the daylight side of the Earth, and times when the on-board astrometry was unreliable (determined by recalculating the astrometric solution using images from the UV/Optical telescope). Details of this filtering were given in paper I. Observations with less than 100 s of PC mode data after such filtering were discarded from the catalog.

Each selected observation was split into snapshots; only snapshots of at least 50 s exposure time[4] (after the above filtering) and at least one X-ray event were included in the catalog. The pointing stability during the snapshot was also determined from the housekeeping data; if the pointing RA (α) or declination (δ) had a standard deviation (from its mean) of more than 25″, the snapshot was discarded. Any observation in which

[1] Excluding non-science observations with target IDs beginning '0006'.

[2] Some of the data taken prior to 2005, i.e. during spacecraft commissioning, have incorrect attitude information as a result of commissioning work. These observations were included in 1SXPS by an oversight.

[3] https://swift.gsfc.nasa.gov/analysis/xrt_swguide_v1_2.pdf

[4] In 1SXPS we required at least 100 s.

Table 1. Summary details of the catalog

Category	Value	Units	Change from 1SXPS
Energy Bands:	Total: $0.3 \leq E \leq 10$	keV	
	Soft: $0.3 \leq E < 1$		
	Medium: $1 \leq E < 2$		
	Hard: $2 \leq E \leq 10$		
Sky Coverage	3,790	square degrees	+99%
Time range	2005 Jan 01 – 2018 August 01		+52%
Usable exposure	266.5	Ms	+81%
Number of observations	127,519		+161%
Number of stacked images	14,545		+98%
Median sensitivity[1] (0.3–10 keV)	1.73×10^{-13}	erg cm^{-2}s^{-1}	-42%
Median source flux (0.3–10 keV)	4.7×10^{-14}	erg cm^{-2}s^{-1}	+50%
Number of detections	1,091,058		+86%
Number of unique sources	206,335		+36%
Number of uncataloged sources[2]	78,100		+14%
Number of variable sources[3]	82,324		+185%

NOTE—[1]The flux at which source detection is 50% complete at the median exposure time. The 2SXPS source detection system is more sensitive than 1SXPS, however the median exposure time in the catalog is also shorter which masks the true sensitivity gain. See Section 7 for more information. The negative sign here shows that the 2SXPS system has a *lower* flux level i.e. improved sensitivity.
[2]Sources without a match within 3-σ in any of the catalogs detailed in Section 5 excluding the 2MASS, USNO-B1 and ALLWISE catalogs, as these have a high rate of spurious matches.
[3]Sources variable with 3-σ confidence in at least one band or hardness ratio.

no snapshots passed these tests was excluded. This resulted in 127,519 observations in the catalog.

As for 1SXPS, we also created 'stacked images' in which all of the observations of a given part of the sky were combined into a single dataset for source detection purposes. This allowed us to maximise the exposure time and hence sensitivity for each given point on the sky. In 1SXPS all images were limited to 1,000×1,000 pixels ($\approx 40' \times 40'$). Since then, Evans et al. (2015) developed tools to allow XRT images to be stacked and analysed by our source-detection tools on an arbitarily-sized grid. For this catalog we set the maximum stacked image size to be 2,300×2,300 pixels ($\approx 90' \times 90'$) which corresponds roughly to a 3×3 grid of XRT pointings. This ensures that the processing time of a given field remains managable, and that the co-ordinate inaccuracy inherent in the tangent-plane projection co-ordinates used for XRT data analysis is negligible. We developed an algorithm to define stacked images such that the minimum number of images is produced necessary to ensure that, for each point of the sky observed by XRT, the maximum possible exposure is reached in at least one stacked image. This yielded 14,628 stacked images. Throughout this work a 'stacked image' is as just defined, while an 'observation' refers to the data organised under a single obsid (which may be comprised of multiple snapshots, usually obtained within a single UT day). The word 'dataset' is used generically to refer to either an observation or stacked image.

The main characteristics of the 2SXPS catalog are given in Table 1, along with a comparison with 1SXPS. In Fig. 1 we show the coverage of 2SXPS. The solid line shows the sky coverage as a function of exposure time (corrected for overlaps). The histograms show the distribution of exposure time in the individual datasets.

3. SOURCE DETECTION

The source detection system employed for 2SXPS was based on that described in paper I with a number of improvements. The algorithm for the detection phase is shown in Fig. 2; steps which are identical to their counterpart in 1SXPS have a blue background, whereas steps which were added or modified for 2SXPS have a yellow background. Here, we briefly summarise the overall algorithm before discussing the modifications in more detail; for an in-depth description of the overall approach see paper I, section 3. We used the same algorithm for both observations and stacked images, except where explicitly noted.

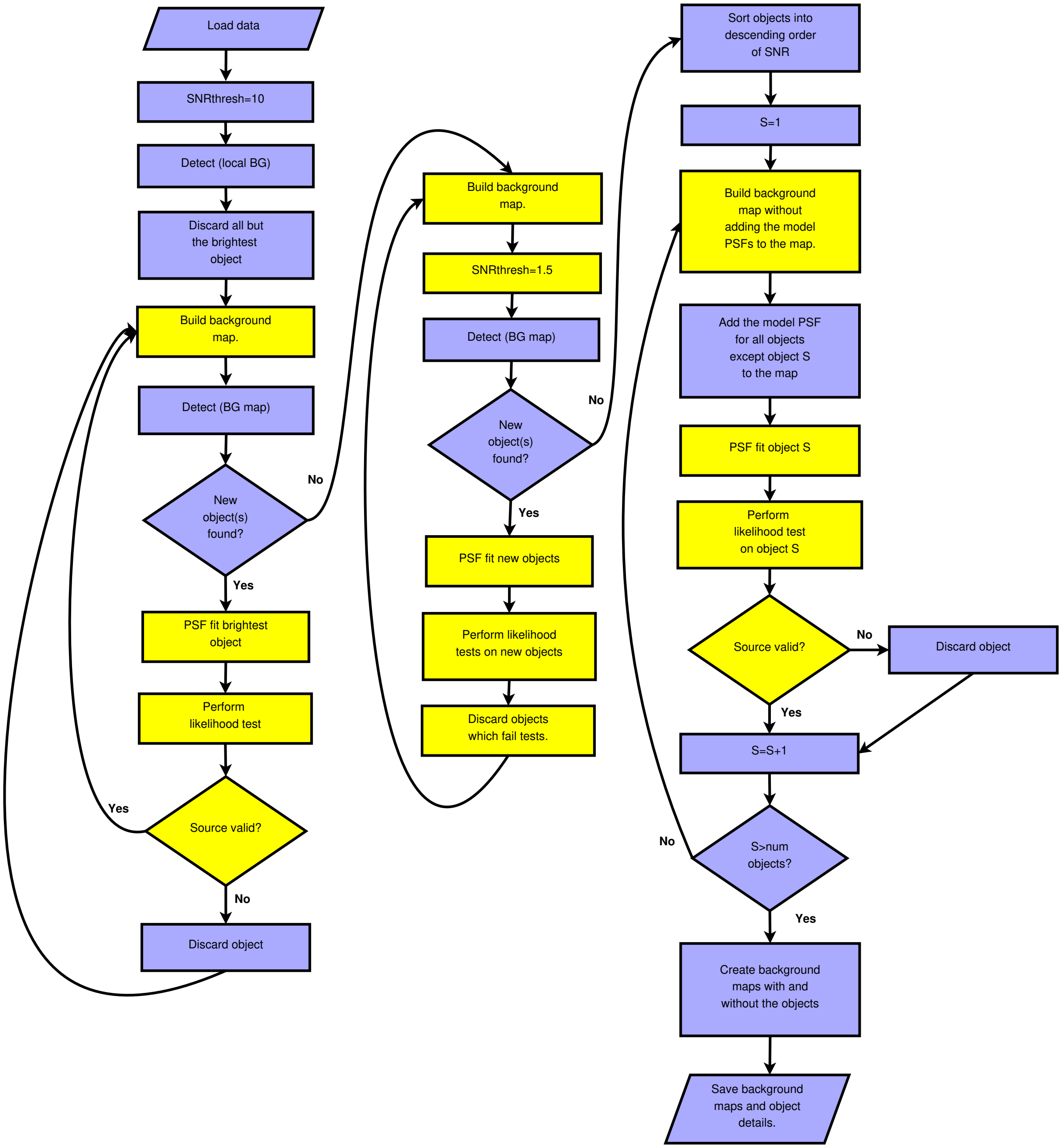

Figure 2. Diagrammatic outline of the source detection process; the overall approach is as in 1SXPS but important changes have been made in the boxes with yellow backgrounds; see text for details of these changes.

To prepare the data for source detection they were split into individual snapshots, for each of which an exposure map and our FITS images were created – one image per energy band in the catalog[5]. For stacked images, the per-snapshot images and exposure maps were shifted onto a common sky coordinate frame (Section 3.1). For each snapshot, the coordinates of the center of the XRT field of view, the window size and the spacecraft roll angle were recorded[6], and any potential sources of stray light were identified and recorded (Section 3.2). The

[5] In paper I, for stacked images of GRB fields we excluded the first snapshot – while the GRB was likely to be very bright and piled up. Due to the improvements made for 2SXPS (Sections 3.3–3.4) this was not necessary for 2SXPS.

[6] So that the background map can be correctly constructed.

per-snapshot images and exposure maps were summed to create a single summed image per band and a single, summed exposure map. The source detection system was then called, once per energy band; it made use of all of the files just described.

Unlike 1SXPS, the detection runs in the four bands were not entirely independent: information about stray light and piled-up sources found in the total band (Sections 3.2 and 3.4) was passed to the other bands. However, no other information was shared between bands at this phase; this contrasts with the approach employed in the 2/3XMM catalogs (Watson et al. 2009; Rosen et al. 2016; Traulsen et al. 2019), where all bands were analysed simultaneously. This is because XRT data must be split into snapshots to calculate the background map, which renders simultaneous fitting across all bands computationally impractical.

The source detection process was a multi-pass process with three distinct phases, shown in the three columns of Fig. 2. It was based on a sliding-cell detection approach combined with PSF fitting. At the start of phase one (left-hand column), an initial sliding-cell detection pass was called for which the background was estimated from a box annulus around the sliding cell. This was used purely to enable the creation of an inital background map (Section 3.2). Thereafter the remainder of phase one and all of phase two followed the same basic repeated pattern: sliding-cell source detection, PSF fitting of the newly-detected source(s), reconstruction of the background map with all detected sources first masked out and then the PSF model of these added into the resultant map.

During the first phase, the signal-to-noise ratio (S/N) threshold for the sliding-cell detection, defined in equation (6) of paper I, was set to 10, and only a single source – that with the highest S/N – was PSF fitted in each iteration. This reduced the number of spurious sources otherwise found around bright sources. Once no S/N>10 sources could be found the second phase (middle column of Fig. 2) began: the S/N threshold was reduced to 1.5 and – because these sources are less likely to yield spurious sources in their wings – all sources detected in each iteration were PSF fitted. In both of these phases, likelihood tests were carried out on each PSF-fitted source (Section 3.5), and sources which did not achieve a status of at least *Poor* were discarded.

Once no more sources were found in the cell-detect pass, the third phase (right-hand column of Fig. 2) was carried out. Here, the PSF fitting was repeated for all sources, using a background map containing the model PSFs of all sources (except that being fitted), allowing a more accurate measurement of each source's properties than was obtained in phases 1-2, where the source list was incomplete and hence the background map inaccurate.

Once this process had been carried out on all datasets, selected observations were manually inspected (Section 3.6), and stray light issues were corrected with source detection repeated if appropriate. Finally, the detections were combined into a unique source list (Section 3.7) and then various source products were created (Section 4).

Two statistics were used in various contexts throughout the fitting process: the C statistic ($\mathcal{C}$, Cash 1979) as modified for use in XSPEC (Arnaud 1996) was the statistic minimised in fitting. A so-called 'likelihood'[7], L, was also calculated at various stages to determine whether one fit was better than another.

$\mathcal{C}$ was defined as:

$$\mathcal{C} = 2\sum_i \left(M_i - D_i + D_i\left[\ln D_i - \ln M_i\right]\right) \tag{1}$$

where M_i is the model-predicted counts in pixel i, and D_i is the actual number of counts in the pixel.

The likelihood reflects the significance of an improvement in fit quality as a result of adding in extra free parameters. Since $\Delta\mathcal{C}$ is distributed as $\Delta\chi^2$, the probability of the improvement arising by chance can be calculated, and the likeliood determined thus:

$$\begin{aligned} L &= -\ln P \\ &= -\ln\left[\Gamma\left(\frac{\Delta\nu}{2}, \frac{\Delta\mathcal{C}}{2}\right)\right] \end{aligned} \tag{2}$$

where $\Delta\mathcal{C}$ and $\Delta\nu$ are the change in fit statistic and degrees of freedom between the two fits respectively and Γ is the incomplete gamma function.

3.1. *Coordinate shifting for stacked images*

For 1SXPS, only 4% of the sky had been observed by overlapping observations that, when stacked, produced an image larger than the 1000×1000 pixel size limit in the standard software tools. Due to new observing modes developed for *Swift*, and their use for observational programs such as the follow up of neutrino detections (Evans et al. 2015; Adrián-Martínez et al. 2016), the S-CUBED survey of the Small Magellanic Cloud (Kennea et al. 2018), the *Swift* Galactic Bulge Survey

[7] The property referred to as a 'likelihood' in paper I and the *XMM* catalogs is not actually a likelihood, or likelihood ratio in the normal statistical sense; it is just the negative of the natural log of a probability. Nonetheless we retain the incorrect use of this term for ease, and consistency with previous work.

(Shaw et al. 2017) and follow-up of gravitational wave events (Evans et al. 2016b), 42% of the stacked fields in this work were larger than this size limit. We therefore created new software to shift the XRT images and exposure maps onto an arbitrary co-ordinate grid. This software made use of the WCSLIB C library[8] (Greisen & Calabretta 2002; Calabretta & Greisen 2002); for each pixel in the original image, WCSLIB was used to convert the (x, y) co-ordinate into (α, δ), and then again to re-convert this into (x, y) in the WCS frame of the stacked image. For the data images, the integer (x, y) positions of each event were converted into floating-point values and randomly[9] positioned within their original pixel. For the exposure maps, the four corners of the original pixel were translated as above to identify the pixel(s) in the stacked image over which the exposure in the original image should be distributed. This exposure was then shared among those pixels according to the fractional overlap. This method was based on the 'area' transform method of the SWIFTXFORM FTOOL.

While this approach allowed arbitrarily-sized stacked images to be created, there were limitations imposed by practical considerations, the chief of which was computational efficiency. The computer resources needed by our source-detection system scale approximately with the number of snapshots, with an additional factor related to overall image size. We therefore imposed a maximum image size of 2,300×2,300 pixels ($\approx 90'\times 90'$) which is sufficient to contain all observations within a standard *Swift*-XRT 7-point automated mosaic, as commonly used the follow up of neutrino triggers, or gamma-ray bursts detected by other satellites.

The data were split into stacked images based on their target IDs: a unique, 8-digit identifier assigned to each target. In principle, all co-pointed observations should have a common target ID, while all observations with a common target ID should be co-pointed. The former constraint was not always true, for operational reasons however this presented no difficulty as co-pointed target IDs were assigned to the same stacked image(s). The latter constraint has occasionally been inadvertantly violated, resulting in a small number of target IDs for which the different observations have disparate pointings. For these cases, the observations were split into co-pointed sets which were then assigned a new (unique) target ID for the purposes of stacked image creation.

[8] http://www.atnf.csiro.au/people/mcalabre/WCS/wcslib/

[9] Randomisation was performed using the GSL_RNG_RANLXD1 random number generator provided by the GNU Scientific Libraries; the seed was based on the computer clock time and process ID of the running task.

In order to ensure that the maximum sky depth was reached for each sky location, target IDs could be assigned to multiple stacked images, and stacked images could overlap. To demonstrate, consider the case of 4 adjacent target IDs along the same line of declination, spaced evenly so as to slightly overlap each other; call these A, B, C, D. These would be split into two stacked images, one comprising A, B and C; the other, fields B, C and D. In this way all of the overlaps (AB, BC and CD) are in at least one stacked image. The sky areas in targets B and C and the overlap BC are in two stacked images, giving duplication of sources, but duplication of sources is an inherent part of the catalog since the observations making up targets A, B, C and D will all have also been analysed separately. The rationalisation of the source lists is described in Section 3.7. In total, 2SXPS contains 34,553 targets contributing to 14,545 stacked images; 7,260 of these target IDs contribute to more than one stacked image. A further 4,022 targets exist which correspond to a unique observation on the sky, and thus are in no stacked image.

3.2. *Background modelling and stray light*

During source detection, the background was repeatedly modelled and the resultant 'background map' was used by the sliding-cell detection and the PSF fitting. The basic approach to background modelling was identical to that in paper I, which was based on that used by the *XMM* (Watson et al. 2009; Rosen et al. 2016) and *ROSAT* (Voges et al. 1999; Boller et al. 2016) catalogs. Sources already detected were masked out, the data were coarsely rebinned and then this rebinned image was interpolated back to each pixel in the image[10]. For any sources which had been PSF fitted in a previous iteration, the PSF model was added to the background map, reducing the likelihood of spurious sources being detected around a bright source and enabling more accurate position determination of nearby sources. This background modelling was conducted for each snapshot separately, since the fields of view in each snapshot do not exactly coalign. The resultant maps were then summed to give a single background map for the dataset.

For 2SXPS we modified the approach from paper I in two ways. First, whereas in 1SXPS the background was rebinned into a 3×3 grid, in 2SXPS for observations longer than 2 ks (i.e. with a better-sampled background) a 5×5 grid was used, enabling locally-elevated backgrounds, for example due to diffuse emission, to be

[10] For *XMM* and *ROSAT* this last stage involved spline fitting not interpolation. The lower, less spatially-variable background of XRT is better handled by interpolation.

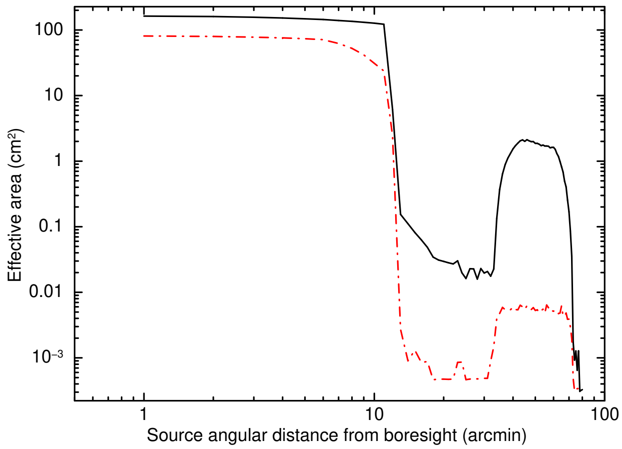

Figure 3. The effective area of the XRT as a function of off-axis angle, derived from ray-tracing simulations. The focussed field of view is 12.3′ in radius. The solid black curve shows the total effective area of the entire CCD. The broken red line shows the effective area over a CCD area equivalent to the half-energy-width of the PSF (18″ diameter). The two curves are more disparate for sources outside the nominal field of view, because the X-rays for these are spread out over a much larger area of the CCD in large rings, whereas for on-axis sources the counts are focussed into the spot-like point spread function.

better modelled. Second, stray light was included in the background map.

Stray light is an artifact of the Wolter-I optic design (Wolter 1952). X-rays within the telescope's nominal field of view undergo two grazing-incident reflections to focus them on the camera: off the parabolic and then hyperbolic mirror surfaces. X-rays from sources marginally outside of the field of view can also be scattered onto the camera via only a single reflection off the hyperbolic surface. Such X-rays fall in concentric rings on the detector (one ring per mirror shell), referred to as 'stray light'. This effect can be predicted and analytically modelled, as described in detail in Appendix A.

In paper I stray light was handled manually, by eyeballing images, identifying regions affected and flagging sources in those regions. For 2SXPS we developed a new technique to automatically include stray light in the background map, dramatically reducing the number of spurious detections. This consists of two main steps: first identifying sources capable of producing stray light and the datasets in which stray light may be expected; then fitting the stray light in the affected images and adding it to the background model.

3.2.1. *Sources of stray light*

The predicted effective area of the XRT as a function of off-axis angle, derived from ray tracing, is shown in Fig. 3. This agrees with the in-flight measurements of Moretti et al. (2009). Any source ∼35–75′ off axis will produce stray light in the XRT, however for most sources this will be so weak and diffuse as to be irrelevant. In 1SXPS the median 0.3–10 keV background rate was 8.6×10^{-7} ct s^{-1} pixel^{-1}, thus only sources bright enough to produce stray light at around this intensity need be considered. The half-energy-width (HEW) of the XRT PSF is 18″, the red dashed line in Fig. 3 corresponds to this, i.e. a region 45 pixels in area, in which the mean 1SXPS background level would be 3.9×10^{-5} ct s^{-1}. The ratio of on- and off-axis effective areas in such a region width is $\sim 3\times10^{-5}$, and by definition the true source count rate of an on-axis source is double the count rate measured in an HEW region. So a source with an on-axis count rate of 2.7 ct s^{-1} or higher can, when 35–75′ off axis, contribute photons to the XRT at a level similar to the normal background.

We compiled a list of all sources in 1SXPS with a count rate above this level. Using PIMMS and assuming a typical AGN spectrum (a power-law with $\Gamma = 1.7$, n_H=3×10^{20}cm^{-2}) to convert this into expected brightnesses in *XMM* and *ROSAT*, we added to this list any source in the HEASARC X-ray Master catalog[11] above this flux. We also added all the contents of the *INTEGRAL* reference catalog (Ebisawa et al. 2003), queried via HEASARC on 2017 July 1: this catalog contains any sources ever recorded above ∼1 mCrab (≈ 1 ct s^{-1} in XRT) at 3keV. This list was then consolidated to remove duplicates and provided a reference list of possible stray-light sources. When the data were split into snapshots, any source in this list which lay 31′–72.5′ away from the center of the XRT field of view was recorded as a possible source of stray light. Because the field of view can vary by several arc minutes between snapshots, this check was done for each snapshot independently.

3.2.2. *Including stray light in the background maps*

In principle, if we know the position of a source with respect to the XRT boresight and its intrinsic flux, the expected stray light from the source can be calculated analytically, as described in Appendix A, and then added to the background map. In practice this cannot be done for two reasons: first, only the 1SXPS and 2/3XMM sources have positions accurate enough for this to be done 'blind'; second, many of the sources are variable and their intensity at the time of the XRT observations is not known. Additionally, the analytical model is not perfect and sometimes the data were better

[11] https://heasarc.gsfc.nasa.gov/W3Browse/all/xray.html, queried on 2017 July 1.

modelled using a slightly incorrect source position. We therefore fitted the predicted stray light to the image.

We defined three free parameters per stray light source: θ, ϕ, N. The first two represent the source position (as position angle relative to the CCD DETX axis, and angular distance off-axis respectively), the third was its normalization. This fitting is a somewhat involved process due to three chief complications. First, unmasked point sources in the image can dominate the fit resulting in very poor reproduction of the stray light. Conversely, since stray light gives rise to spurious detections, masking out point sources can result in the stray light being entirely masked out and so unfittable. Second, the stray light contribution should be separated from the underlying background, otherwise the rebin/interpolate approach to creating the background map will overestimate the background in regions near to the stray light. Third, the pointing direction can vary by several arc minutes between snapshots, which is sufficient to significantly change the stray light pattern. Due to the inaccuracies in the model (Appendix A) it is not sufficient to identify the position and normalization of the stray-light-causing source in one snapshot and then simply adjust the parameters according to the pointing differences.

The algorithm developed to surmount these issues and provide a model of the stray light is described in Appendix A.1. It was optimised by running it on a series of 1SXPS datasets with and without stray light. Even so, a visual check was made of all possible stray-light fields, as described in Section 3.6.

3.3. *PSF model and fitting*

The PSF-fitting of new sources proceeded largely as in paper I: a circular region was identified, centered on the cell-detect position and with a radius depending on the source S/N, the source position (x, y) and normalization were identified as free parameters, and $\mathcal{C}$ (Equation 1) was minimised. Minimization was carried out using the MnMinimize class in the minuit2 C++ libraries[12]. The position uncertainty was found for each axis independently, by stepping the position in that axis and refitting (while keeping the test position frozen); where $\mathcal{C}$ increased by one from the best-fitting value gave the 1-σ confidence interval on the position in that axis. Very occasionally in 2SXPS, the position error could not be found in this way: MnMinimize failed to return a valid fit while the source position was being stepped around. In this case the radial position error of the source was set to $\frac{11.301''}{\sqrt{N}}$ (90% confidence), where N is the number of counts in the PSF fit; this relationship giving the best-fit to the 2SXPS position errors determined succesfully by $\mathcal{C}$ stepping. There were also some cases where position errors were found, but were much smaller for the number of counts than was typical of the catalog. Such values may indicate that the $\mathcal{C}$ stepping encountered difficulties, but equally there are cases (e.g. crowded fields) where $\mathcal{C}$ can vary sharply with position. We did not alter these small values.

A few changes from the paper I approach need to be noted. First, sources with a S/N $\geq$60 from the cell-detect phase were fitted over a region with a radius of 40 pixels (in paper I everything with S/N$\geq$ 40 had a radius of 30 pixels). Second, if the position returned by the PSF fit had moved from the input position by more than 50% of this radius, the fit was repeated using a new region centered on the new position. A source could be refitted in this way no more than 5 times (to prevent infinite loops if a degenerate position solution was found). This was beneficial because, for very piled up sources where the PSF core has no counts in it (see Section 3.4), the true source position could lie outside of the initial PSF-fit region. A third change relates to the way pile up was handled, and will be discussed in the next section.

As well as these procedural changes, we considered the shape of the XRT PSF. Within the *Swift* software and calibration database (CALDB) the PSF is defined as the combination of a Gaussian and King function:

$$P(r) = N e^{\frac{r^2}{2\sigma^2}} + (1-N)\left[1 + \left(\frac{r}{r_c}\right)^2\right]^{-\beta} \tag{3}$$

where N is a normalization, r the radius at which the PSF is evaluated, and σ, r_c and β the parameters controlling the shape of the Gaussian and King components. Moretti et al. (2005) calibrated this in-flight and deemed that only the King-function component (the second part) was necessary, i.e. $N = 0$. While this proves a good description of most sources, we have found that for bright objects the outlying wings of the PSF appear to be underpredicted by this model, consistent with fig. 5 of Moretti et al. (2005). This results in the background map around bright sources being too low and spurious sources being detected around bright objects. In paper I we handled this by defining a 'blind spot' around bright sources, in which detections were discarded as likely duplicates of the central object. This approach is less than ideal, as real objects do appear near bright ones. For this paper, we therefore attempted a recalibration of the PSF, in order to better model the wings.

[12] http://project-mathlibs.web.cern.ch/project-mathlibs/sw/Minuit2/html/index.html

Table 2. The PSF parameters derived for and used in 2SXPS. The PSF is defined in Equation 3.

Parameter	Value
N	0.080
σ	3.119 pixels (=7.351″)
r_c	1.597 pixels (=3.764″)
β	1.282

This work is described in Appendix B, and produced the PSF parameters shown in Table 2. This PSF was used throughout this work, and will replace the existing PSF definition in a future CALDB release. This dramatically reduced the number of spurious detections around bright sources. In paper I (Appendix A), we derived a function to model the 'spokes' in the PSF (the shadows of the mirror support structure): this is a function of PSF radius and azimuthal angle and the original PSF model is multiplied by this function. This function is not affected by the new PSF definition and was used as in paper I.

3.4. *Pile up*

Pile up is a phenomenon affecting photon-counting detectors such as the XRT. It occurs when multiple photons impact the same or adjacent CCD pixels within a single exposure frame, and on read-out the charge thus liberated is interpreted as arising from a single photon. Since this is a stochastic process some fraction of the events from any source will be affected by pile up, however this fraction only becomes significant at moderate source intensities: in XRT PC mode typically pile-up starts to become a factor for sources around 0.6 ct s^{-1}. Pile up is initially manifested by the core of the PSF being slightly suppressed compared to the wings, and the source spectrum being artificially hardened. A second factor is so-called *grade migration*: events are assigned a grade based on how many adjacent pixels are affected by the cloud of charge liberated by the incident X-ray. In the case of pile up, separate X-rays incident on adjacent pixels will be erroneously recorded as a single event covering both pixels. Once pile up becomes severe, this causes events to have invalid grades[13] and thus be rejected, resulting in an apparent 'hole' in the core of the PSF; an example of such a source is shown in Fig. 4.

[13] That is, grades above 12 for PC mode, see https://www.swift.ac.uk/analysis/files/xrt_swguide_v1_2.pdf.

Evans et al. (2009) developed a series of discrete PSF profiles whereby a Gaussian component was subtracted from a King component, which approximately described the PSF at increasingly degrees of pile up. In paper I, each of these profiles in turn was applied to a source and the most appropriate profile was determined based on the fit statistic. Since we have redefined the PSF for this work (Section 3.3) these old profiles can no longer be used; and because the new PSF has both King and Gaussian components, the addition of a third element would also be incompatible with the existing CALDB and XRTDAS software. However, pile up can be very accurately modelled simply by multiplying the PSF by an analytical multiplicative loss function. This function was original used by Popp et al. (2000) to describe the spectral energy redistribution of the *XMM* EPIC-pn camera but works well in our context. It depends only on radius r and is given by:

$$f(r) = S + B\left(\frac{r}{l}\right)^c :: (r < l) \tag{4}$$

$$f(r) = 1 - Ae^{-\frac{r-l}{\tau}} :: (r \geq l) \tag{5}$$

where

$$B = \frac{l(1-S)}{l+c\tau} \tag{6}$$

$$A = 1 - S - B \tag{7}$$

provided $S < 1$; otherwise $A = B = 0$ and the function has no effect. Thus, pile up can be modelled by the addition of the following four free parameters. S, which can be in the range $[0, 1)$ determines the overall depth of the loss function, for $S = 0$ there is a hole in the center of the PSF, at $S = 1$ pile up has no effect. l, which was limited to [0.1–50] controls the overall scale of the loss function, and the transition from the core to the wings of the PSF. c, which we restricted to $[0.1, 10]$ affects the steepness with which the loss-function changes in the PSF core, and is complemented by τ which could cover $[0.1, 200]$ and controls the loss-function out in the PSF wings. With the exception of S we had no *a priori* expectations of what ranges the parameters should cover, and the above ranges were taken as those which (from CCD simulation work, Beardmore et al. in prep) could be deemed reasonable.

When performing a PSF fit to a source, the fit was originally carried out using the new PSF model (Equation 3) with no loss function. If the source had a S/N from the cell-detect pass of at least five, a second PSF fit was performed, this time with the loss function included, and hence four extra free parameters. As with the original fit, if this moved the position significantly, the fit

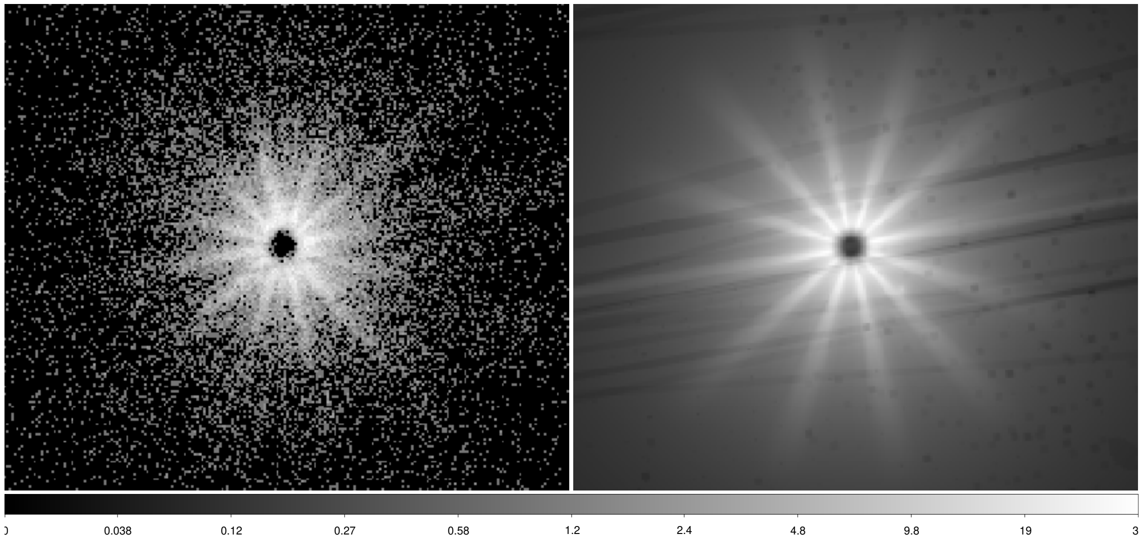

Figure 4. The effect and successful modelling of pile up. *Left*: An extremely piled up source, 4U 1820−30, in which the centre of the PSF contain no counts due to grade migration (see text). *Right*: The background map of this dataset containing the fitted PSF model of the source, showing that pile up has been well reproduced.

was repeated with a new region centered on this position: see Section 3.3. The likelihood value relating to the new fit was calculated using Equation 2, where $\Delta\mathcal{C}$ was the difference between the with/out loss-function fits, and $\Delta\nu = 4$ (the loss function parameters). If $L > 10$ the source was deemed to be piled up, and the results of the fit with a loss-function were taken as the source parameters.

If a source is affected by pile up, the PSF shape will be affected in all energy bands (although not necessarily to the same degree as pile up causes soft events to migrate to the harder energy bands), regardless of the brightness in that band, which can cause problems for the algorithm as described above. Consider for example, a very piled-up, very absorbed source. There may be only a small number of events in the soft band, thus the source will have a low S/N and so not meet the criteria for the piled-up fit to be performed. But, those few counts will nonetheless show a hole in the center of the PSF and the non-piled-up fit will give an inaccurate position. In order to properly handle such events, a list of sources found to be piled up in the total band was supplied to the processing for the other energy bands. Any source found in those bands which lay within 20 pixels of a piled up source (50 pixels if $S < 0.1$) was assumed to be the piled up source, and thus the loss-function fit was performed regardless of the S/N; the L threshold required for such sources to be recorded as piled up in the sub-bands was reduced to 2.5. Despite this, there were still cases where pile up was not properly identified in the sub-bands, and instead multiple faint, non-piled up sources were reported. These were identified and handled during the creation of the unique source list (Section 3.7).

For all sources for which the loss function was fitted, regardless of whether it was accepted as necessary, the best-fitting loss function parameters were included in the catalog, along with $\mathcal{C}$ with and without the loss function and a note of whether the preferred fit was that with or without pile up.

As can be seen from the above description, in our software the loss function was applied to the PSF, i.e. it affects only the events expected from the source. In reality, the situation is more complex since there will also be background events present, and pile up is related only to the event rate, not the event origin; that is, the background should also be suppressed by pile up; but the loss function definition does not account for this. In fact, this issue is generally irrelevant because the source is, by definition, extremely bright and the background negligible in comparison. The exception is for cases where $S \to 0$, giving a hole in the center of the PSF as all events are migrated to unfeasible grades or energies. In reality, there will be no events in the CCD center because of pile up, however our PSF model will contain no *source* counts, but background events are still present. Since the hole is symmetrical, and the fit will be dominated by those regions where source counts are present, this problem can be discounted.

3.5. *Likelihood tests and flags*

In paper I, we determined whether a detection corresponded to a real source by means of a ‘likelihood’ value (hereafter $L_{\rm src}$) as defined in Equation 2, where

$\Delta\mathcal{C}=\mathcal{C}_{\rm nosrc}-\mathcal{C}_{\rm best}$: here $\mathcal{C}_{\rm best}$ is the fit statistic of the PSF fit, and $\mathcal{C}_{\rm nosrc}$ is the C-stat value obtained comparing the background map with the data over the source PSF fit region, i.e. the C-stat obtained if there were no source present. For non-piled-up sources, $\Delta\nu$ is 3, for piled-up sources it is 7 (i.e. the number of free parameters in the PSF fit). Sources were assigned a quality flag based on $L_{\rm src}$ (calibrated via simulations, see Section 7). The limitation of this statistic is that it can have a high value for reasons other than the presence of a point source: for example diffuse emission or imperfectly-modelled stray light (Section 3.2.1) may be 'better fitted' with a PSF-like distribution of counts than only with the underlying background, despite there being no point source present in reality. We therefore introduced an extra test for 2SXPS, to supplement $L_{\rm src}$. A model was fitted to reproduce a homogeneous elevation in count rate in the fit region. This model had a single free parameter: the normalization. The C-stat for this model ($\mathcal{C}_{\rm flat}$) was recorded and $L_{\rm flat}$ calculated via Equation 2, comparing $\mathcal{C}_{\rm flat}$ with $\mathcal{C}_{\rm best}$. Low values of $L_{\rm flat}$ indicated that the PSF-like count distribution offered little improvement over a homoegeneous disitrbution, i.e. the 'detection' was unlikely to be a point source.

As in 1SXPS, we defined three possible source flags: *Good*, *Reasonable* and *Poor*[14], and like 1SXPS, these were defined such that the spurious source contamination level was 0.3%, 1% and ≤10% in the *Good*, *Good+Reasonable* and full catalog samples respectively. However, this time when determining the source flag both $L_{\rm src}$ and $L_{\rm flat}$ were taken into account. The relationship between these likelihoods and source flag depends on the exposure time. In paper I we determined this relationship based on the exposure in the image, which, because vignetting in XRT is modest, was a viable approach. Due to the larger stacked images in 2SXPS this is no longer viable as exposure can vary dramatically across the image due to the varying number of overlapping observations. The exposure time used in flag determination in this work was thus the exposure time at the location of the source. Additionally, the dependence on exposure time is really a proxy for dependence on the background level. The L thresholds in paper I were determined using simulated total-band images, and thus were likely over-conservative for the soft, medium and hard bands, in which the background level is naturally lower. For this work we instead determined the mean background levels from 1SXPS in each of the energy bands as a factor of that in the total energy band. When calculating the exposure to use in determination of a source's flag, the actual mean exposure time at the source position was multiplied by this factor.

[14] 1SXPS also contained 'bad' sources with a very low likelihood of being real. We dropped this for 2SXPS.

Table 3. The threshold likelihood values for the different detection flags; both likelihoods must be above these for a source to be given the described flag.

Flag	Exposure[1] range	$L_{\rm src}$	$L_{\rm flat}$
Good	$E \geq 1000$ s	$18.293E^{-0.0607}$	4
	$300 \leq E < 1000$ s	— " —	0
	$E < 300$ s	14.8	0
Reasonable	$E \geq 1000$ s	$14.788E^{-0.0562}$	6
	$300 \leq E < 1000$ s	— " —	0
	$E < 300$ s	12.7	0
Poor	$E \geq 40000$ s	$7.7873E^{-0.0433}$	6
	$300 \leq E < 40000$ s	— " —	0
	$E < 300$ s	6.4	0

NOTE—[1] E = exposure at the source position, scaled by the background in the given band relative to the total band.

The relationship between L values, exposure and assigned flag was calibrated via the simulations described in Section 7. As in paper I, we found that the threshold L values depended on exposure time, as shown in Table 3. Sources flagged *Good* by their likelihood values were downgraded to *Reasonable* if they lay within 30 pixels (71″) of fitted stray light emission, or if the mean background in the source region was above 10^{-3} ct s^{-1}; this latter case indicating that the detection was likely to have arisen in an area heavily affected by the PSF wings of a bright source. Such detections can be real sources, but the contamination rate in these cases will be higher than in the simulation used to calibrate flag settings; we demoted such sources to keep the *Good* sample as pure as possible.

Within the database table, the flags are stored as integer values: 0, 1, and 2, corresponding to *Good*, *Reasonable* and *Poor* respectively. These flags could be increased to indicate concerns regarding the source. The extra values are bit-wise flags, described in Table 4. So, for example, a source with a flag value of 5 would mean that the source is *Reasonable* (based on it likelihood values) but corresponds to a position covered by a known extended source; thus it may be a point source within the extended emission, or it may be a spurious event arising due to the extended emission.

Table 4. Definition of the bits in the detection flag that were set to indicate a potential problem.

Bit	Value	Meaning
2	4	Source is within the extent of a known extended source.
3	8	Source likely a badly-fitted piled-up source[1].
4	16	Position matched area flagged by visual screening[2].

NOTE—[1] See Section 3.7.
[2] See Section 3.6.

As well as the detection flag, three other flags were created for each source. 'StrayLightWarning' was 0 or 1, indicating whether the source had been flagged as being affected by stray light (defined above). 'NearBrightSourceWarning', indicates whether the mean background level at the source location was high and so the source may be spurious due to a nearby bright object (see above). A value of 0 indicates that this warning is not set, and 1 indicates that it is. A value of 2 can also be given for sources detected in stacked images. In these cases, if there is a variable source which was briefly bright and has been observed many times, the PSF wings in the stacked image will have a low overall count-*rate*, and the time-averaged PSF model may underestimate the PSF wings. So for any source detected in a stacked image, in which the background rate is high (i.e. above 10^{-3} ct s^{-1} pix^{-1} as above) in any individual observation of the source's location, the 'NearBrightSourceWarning' is set to 2; the flag associated with that detection is also downgraded from *Good* to *Reasonable* if it was the former. Another flag, 'OpticalLoadingWarning', indicates whether the source was potentially affected by optical loading; that is, whether its position matched a known optical source bright enough to deposit sufficient energy onto the XRT to masquerade as X-rays[15]. If no such optical source was found, this flag was set to 0, otherwise its value indicates how many magnitudes brighter the optical source is than the magnitute at which optical loading is first expected to be a factor. Optical loading is discussed in more detail in paper I, section 3.4.

3.6. *Visual screening and flagging of datasets*

After source detection had been completed in all four energy bands, a dataset was flagged as needing manual examination if any of the following criteria were satisfied:

1. A possible source of stray light had been found (regardless of whether it was deemed necessary in the fitting).
2. The median distance between detections in any given band was $< 80''$.
3. The dataset corresponded to an observation in 1SXPS which had a non-zero flag after manual screening in that catalog.
4. The dataset was a stacked image, for which one of the component observations satisfied criteria 2 or 3 above.

The first criterion required us to verify that the stray light modelling was at least adequate in all cases. Criterion 2 was specified because a high density of observations either indicated a genuinely dense field (such as the core of M31), or the presence of an artifact that gave rise to multiple spurious detections such as diffuse emission or unmodelled stray light. Criterion 3 is self-explanatory, and criterion 4 ensured that any stacked image containing a potentially contaminated observation was also checked. In total 13,825 datasets (out of 142,064) were identified in this way. For each of these two questions were addressed: whether there was any diffuse emission present in the image and whether stray light was handled adequately. In the former case, if diffuse emission was identified in the image we created circular or elliptical region(s) to cover the emission. All sources lying inside this region were flagged (bit 4 of their flag set, see Section 3.5 and Table 4). If this emission was astrophysical (i.e. not arising due to instrumental effects such or bright Earth contamination) then the region was also applied to all other observations sharing a target ID with the screened image, and to any stacked image the observation in question contributed.

For fields where stray light had been deemed unnecessary by the fit, we confirmed that it was indeed absent. For observations where stray light had been fitted, the stray light model was compared with the image to confirm that stray light was indeed present, the model gave a reasonably accurate reconstruction of the stray light and that any detection within the stray light was appropriately flagged. In the event that stray light was present but not fitted, or was fitted but badly, we manually generated stray light images for trial positions of the causal source until a reasonable reproduction of the observed stray light was obtained. The observation was then reanalysed, with the manually-determined position

[15] See https://www.swift.ac.uk/analysis/xrt/optical_loading.php.

Table 5. Definition of the field flag, assigned to each dataset.

Bit	Value	Meaning
0	1	Stray light was present, and fitted.
1	2	Diffuse emission identified.
2	4	Stray light badly/not fitted.
3	8	Bright source fitting issues[1]

Note—[1] i.e. the field contained a source that was heavily piled up in one band, but not fitted as such in another band. See Section 3.7 for details.

provided as the starting point for the stray light fit. The field was then reinspected to confirm that the stray light was now handled. Fields where stray light was present but had not been modelled because the causal source was not found in the catalog look up, were selected for screening as they met criterion 2 above, they were handled in the same way as sources where the stray light was badly fitted. In some cases even after refitting, the model fit was clearly imperfect (for example, the curvature of the rings was not quite right, or an extra ring was modelled which was not seen in the data); however, provided that the model had been able to suppress spurious detections, or at least ensure that such detections were flagged, it was accepted as 'good enough'. In a small number of cases (104) even after this iteration an acceptable reproduction of the stray light could not be obtained. For such fields bit two of the 'field flag' for the affected dataset was set, as described below.

For observations where stray light had been modelled, but visual inspection showed that there was in fact no such contamination the observation was reanalysed with no stray light fitted. If this reanalysis resulted in a median inter-source distance of $< 80''$ we reinspected the field to confirm whether our decision to remove the stray light model had been erroneous (in which case it was reinstated). If an observation had to be reanalysed as a result of the stray light screening, all stacked images to which that observation contributed were also reanalysed.

Once visual screening was complete, each dataset was assigned a flag, referred to as the 'field flag' in the catalog tables. This is a bitwise flag, and the different flags are defined in Table 5.

3.7. *Construction of the unique source list*

The rationalization of detections into a unique list of sources was a two-step process: first the detections from the different bands within each dataset were combined into a single source list per observation. For this step, the astrometric uncertainty on the XRT pointing could be neglected since it was the same for each band. The second step was to combine the outputs of step one from each dataset into a unique list of sources; for this of course the uncertainty in the relative astrometry of the different observations had to be accounted for.

For step one, two detections in different bands were considered the same source if their positions agreed to within 3 σ or they were within 10 pixels ($23.6''$) of each other. The latter clause arose because our simulations showed a large tail on the position reconstruction error. This differs notably from 1SXPS where the match radius was simply a function of source brightness. The reference position for the source was taken from the *Good* or *Reasonable* detection with the smallest position error or, if all detections were *Poor*, the detection with the highest S/N. In the event of a detection in one band matching multiple detections from other bands, it was assigned to that to which it was closest.

The exception to this was for piled up sources. As discussed in Section 3.4, it is possible for a heavily piled-up source not to be identified as such in the sub-band images, giving rise to multiple spurious detections. To ameliorate this issue the detection-matching process was carried out for piled-up sources first. For each piled up detection, d_i, in a given band, b, a counterpart was sought in each other energy band with a localisation within 5 pixels ($11.8''$) and an S value within 0.5 of that of d_i. If such a detection was found in band b it was associated with d_i; if not, then the pile up profile of source d_i had not been fitted properly in band b, therefore all sources in that band within 100 pixels ($236''$) of d_i were assumed to be spurious detections of d_i. Such detections had bit 3 of their flag set (Table 4).

Once the unique list of sources per dataset had been determined (in step one), the absolute astrometry of the dataset was calculated by aligning these sources with the 2MASS catalog (see paper I, section 3.7 for technical details). If this process was successful, the corrected positions were reported, but they were only used if the uncertainty on the 2MASS-derived astrometric solution was smaller than that associated with the XRT star tracker astrometry ($3.5''$ at 90% confidence).

For step 2, the unique source lists from each dataset were compared, and objects were considered the same source if their position – including astrometric uncertainty – agreed at the 5-σ[16] level. The final detection flag assigned to each source and band was the best flag from all the individual detections in that band; the overall detection flag and S/N for each source was the best obtained from all detections and bands. The stray light and optical loading warnings for each source were set to the worst values from the individual detections.

In a small number of cases multiple detections of the same source were erroneously recorded as different sources, as their positions in the different detections differed by more than 5-σ, suggesting either a high proper motion, or that the position errors have a larger tail than would be expected from pure Rayleigh statistics. The latter case will occur if, for example, the astrometric solution related to a field has degenerate solutions as can occur if (for example) the number of reference stars is low. We therefore identified any sources which were within $20''$ of each other and not identified in the same dataset, and marked them as potential aliases of each other. 1,735 sources were identified in this way. Not all of these are aliases: some will be spurious events around a bright source, and some genuinely nearby but distinct sources. However, these possible aliases are marked to allow users to investigate more closely if they desire.

4. SOURCE PRODUCTS

For each source we determined the count rate for each energy band and observation covering the source location, regardless of whether it was detected in that dataset. We measured these rates both averaged over the observation and for each individual snapshot. A circular region centered on the source position was used, with the radius set to that used when the source was PSF fitted, or 12 pixels ($28.2''$) if the source was not detected in the dataset under consideration. The total counts in that region, C, was measured from the image, and the expected number of background counts B was taken from the final background map for the observation/snapshot. If the source had been detected in the observation in question, the PSF model for the source was first subtracted from the background map. If $C - B > 100$ or either $C > 1000$ or $B > 1000$ standard frequentist statistics were used to determine the number of source counts and its error; otherwise the Bayesian approach of Kraft et al. (1991) was used. As in paper I, we calculated the 1-σ confidence interval on all count rates. However, in addition for this work we calculated the 3-σ interval for all observations and bands in which the source had not been detected, and for all snapshots. If the 3-σ lower limit was 0, the source was flagged as undetected in this dataset, and the 3-σ upper limit was recorded as well as the 1-σ confidence interval. Note that a source which was not found by the source detection process in a given dataset can nonetheless be reported as detected in the same dataset by this 'retrospective' count rate calculation approach; this is because the source detection is a blind process, whereas retrospective count rate measurement is predicated on the knowledge that there *is* a source at that location, which makes it more sensitive (i.e. one does not need to allow for the large number of trial positions). When accessing the source light curves via the 2SXPS website, users can choose whether to define a datapoint as a 'detection' based on the blind search or the retrospective analysis, and whether to retrieve 3-σ upper limits or 1-σ confidence intervals for non detections, giving greater control than was possible in 1SXPS. In addition to these time-resolved count rates, a single mean count rate per energy band was determined by summing C and B from all the individual observations. The peak rate in each band was also recorded, determined as the individual per-observation or per-snapshot rate with the highest 1-σ lower limit.

All count rates above were corrected for vignetting, pile up and bad columns or pixels on the CCD. This was done by summing the fitted PSF model (with pile up, if appropriate) multiplied by the exposure map over the circular extraction regions, then also integrating the theoretical PSF to a radius of 150 pixels[17] multiplied by the peak on-axis exposure time. The ratio of these gives the correction factor by which the count rate and error were multiplied. Note that, for large stacked images the fractional exposure towards the edges can be very small compared to the peak exposure time, giving very large corrections. When calculating the mean count rate the correction factor was calculated as $\sum_i (C_i F_i) / \sum_i C_i$, where C_i is just C measured from dataset i, and F_i is the correction factor in that observation.

As well as light curves, two hardness ratios were created for each source, for each snapshot and observation and an overall ratio. These ratios were defined as in paper I:

$$\mathrm{HR1} = (M - S)/(M + S) \tag{8}$$

$$\mathrm{HR2} = (H - M)/(H + M) \tag{9}$$

[16] Specifically at the probability associated with a Gaussian 5-σ confidence. Since the radial position errors should follow a Rayleigh distribution, this level was determined based on Rayleigh, not Gaussian, statistics.

[17] i.e. effectively infinity.

where S, M, H refer to the background-subtracted soft, medium and hard bands respectively (the bands were defined in Table 1). If both bands in the hardness ratio contained > 100 counts and had S/N> 2 then the ratios were calculated using the above equations, with the errors on H, S and M taken as $\sqrt{\{H, M, S\}}$ respectively and propagated through Equations 8 and 9. For fainter sources we used the Bayesian method of Park et al. (2006), where we used the effective area option in their code to include the count rate correction factors in the calculation.

For a small number of datasets with short exposures, there were no events in one or more of the sub-bands, in which case the HRs could not be determined.

For each energy band and hardness ratio we also quantified source variability. This was done by creating per-snapshot and per-observation light curves from the count rate and hardness ratios calculated as above; the 1-σ confidence intervals were used for all bins. The Pearson's χ^2 (Pearson 1900) was then calculated as in paper I, where the model was that of constant flux at the mean level, and from this the probability that the source was constant was determined (see paper I, section 4.1 for details).

Note that, for all data products, we used only the PC mode data included in the catalog. Many of the sources have also been covered by WT mode observations. However, these contain only 1-D spatial information and so are only appropriate for bright sources: with the majority of 2SXPS sources being serendipitous, the WT data will be contaminated by the other sources in the field[18].

As noted above, a very small fraction ($< 0.8\%$) of the sources in the catalog are potential aliases of other sources; in these cases the light curve will contain a mixture of correct source count rates and erroneous measurements or upper limit (the latter in the case that the detections of the source were classed under its alias). The 2SXPS website (Section 6) allows the light curves of aliases to be combined in order to give the correct data.

4.1. *Spectral information and flux conversions*

Spectral and flux information was determined for every source. The approach is summarised briefly here, for full details see section 4.2 of paper I. These values were determined for two spectral models: an absorbed power-law, and an absorbed optically thin plasma model (APEC Smith et al. 2001); absorption was calculated using the TBABS model (Wilms et al. 2000). Flux conversions and (where appropriate) spectral properties were derived using XSPEC. Up to three methods were used to determine the spectral details for each model.

The first method was applied to every source. We assumed standard emitting models: a power-law with photon index 1.7 and an APEC with a plasma temperature of 1 keV, and fixed the absorption column to the Galactic value along the line of sight to the source, from Willingale et al. (2013). The second method was attempted for every source. For this we simulated spectra in XSPEC to produce a look-up table of the spectral parameters (i.e. absorption column and either power-law photon index or APEC temperature) as a function of (HR1,HR2). For each source whose time-averaged (HR1,HR2) values were consistent with those producible by such a spectrum, we interpolated on this grid to determine the spectral parameters (and uncertainties), and hence also the energy conversion factor[19]. The third method was only carried out for sources from which more than 50 net counts were detected, and which were detected (either in the blind searching or the restrospective count rate determination) in at least one single-observation dataset. For these, a spectrum was constructed using the tools from Evans et al. (2009), combining only those individual datasets in which the source was detected (again, via either definition) – this is to avoid diluting the S/N in the spectrum by including periods of background-only data. In this case the spectral models were fitted to the extracted spectra to give the best-fitting parameters. Fitting was carried out on the unbinned data, minimising the $\mathcal{W}$ statistic in XSPEC; a Churazov-weighted χ^2 (Churazov et al. 1996) was then calculated to give a goodness-of-fit indication[20]. The energy conversion factors (ECF) for the source (i.e. conversion from count-rate to flux) were then derived from these spectra, but the fluxes in the catalogue were found by multiplying the mean total-band count-rate by the ECF. That is, by limiting ourselves to datasets where the source was detected we have not biased the fluxes in the catalog.

The only deviation of this approach from the paper I method affected the second method (HR interpolation). In paper I we created a single set of look-up tables for each of the two spectral models. However, on 2007 August 30 the CCD substrate voltage was changed from 0V to 6V. This has a small effect on the spectral calibration of the instrument, so for this work we created separate look-up tables for the two substrate voltage settings. We

[18] Normally WT mode is only used for bright sources, where the number of photons from field sources is negligible compared to those from the source.

[19] i.e. the conversion from detected counts, to source flux.

[20] Note that this χ^2 cannot be used to calculate the null hypothesis probability.

chose which table to use based on whether the mean arrival time for photons from the source occured before or after the voltage change.

For the APEC spectral model, there is a small region of (HR1, HR2) space which would be occupied by sources with very high absorption columns ($> 10^{22}$ cm^{-2}) and typically low (< 1 keV) plasma temperatures; for such sources the predicted counts to *unabsorbed* flux conversion is very high (due to a low predicted count rate, but high unabsorbed flux). There is a small number of sources, <1,000, in 2SXPS which thus contain unrealistically high unabsorbed flux values, based on the interpolated APEC spectrum; such values should be treated with caution, and are more likely to indicate that the true source spectrum is not an absorbed APEC. The *observed* fluxes for these objects are not unrealistic, since these have, like the count rate, been suppressed by the high absorption.

5. CROSS CORRELATION WITH OTHER CATALOGS

We cross-correlated the 2SXPS catalog with a range of other catalogs, using the same approach as paper I (section 4.3), i.e. identifying all sources in those catalogs with positions agreeing with the 2SXPS position at the 99.7% confidence level (using Rayleigh statistics, accouting for the uncertainty in the 2SXPS and external catalogs). Unlike paper I we chose not to correlate against the dynamic catalogs of SIMBAD and NED (links to perform such a search are provided on the 2SXPS website), but we added correlations with ALLWISE and 1SXPS. For the other catalogs we used updated versions if they existed: the list of catalogs and number of matches are given in Table 6. As for 1SXPS, we estimated the rate of spurious correlations by randomly shifting the 2SXPS positions by 1–2′ and repeating the correlation: the number of matches found in this second pass is also shown in Table 6.

6. CATALOG CHARACTERISTICS, ACCESS AND CONTENTS

2SXPS contains 206,335 unique sources, with a total of 1.1 million blind detections across all four energy bands[21]. The median 0.3–10 keV flux[22] is 4.7×10^{-14} erg cm^{-2} s^{-1}. The observations in the catalog contain a total of 267 Ms, with a unique sky coverage of 3,790 deg^2. This is nearly twice as much sky area as was covered by 1SXPS, 3.5 times the area covered by 3XMM-DR8, and 6.8 times that in CSC 2.0[23]. There are 82,324 variable sources in the catalog. Despite the lower effective area of XRT compared to the *XMM* instruments, the median source flux is only a factor of two higher than in 3XMM-DR8, likely due to the lower background level in XRT caused by its low-Earth orbit. The median source flux is actually higher than in 1SXPS, despite the fact that our improved source detection system is actually more sensitive (Section 7); this is because the combination of our different data selection criteria and the evolution of the science *Swift* carries out, results in a mean observation exposure time in 2SXPS of 2063 s, compared to 3007 s in 1SXPS.

The catalog can be queried or downloaded via a bespoke website at: **https://www.swift.ac.uk/2SXPS**. Four tables are available for download, containing the sources and their properties, individual detection details, details of the datasets in the catalog, and details of the external catalog cross-correlation. The contents of these tables are described in Appendix C, Tables 8–11. The main table, detailing the unique sources, is also is available through Vizier, as catalog IX/58, and will be made available through HEASARC.

The source and datasets tables can be queried via the above website, either using a simple cone search or using detailed filtering on any/all of the table properties. Web pages exist for each source and dataset, giving access to all products. An upper limit service is also provided. Full documentation is on the website.

As for 1SXPS we have defined a set of filters defining a 'clean' sample, and additionally for 2SXPS an 'ultra-clean' sample. Cone searches on the website can be restricted to these subsamples. Clean sources are those with a best detection flag of 0 or 1 (i.e. *Good* or *Reasonable* with no other warning bits set); OpticalLoadingWarning, StrayLightWarning and NearBrightSourceWarning all unset; and a field flag or 0 or 1 (see Table 5). Ultra-clean sources are are subset of the clean sources, with detection and field flags of 0. There are 146,768 clean sources and 132,287 ultra-clean sources in 2SXPS.

7. COMPLETENESS, CONTAMINATION AND ACCURACY

As for paper I, we used simulations to calibrate the likelihood thresholds and explore the performance of our source detection software. We used the background maps (minus source models) from 1SXPS as the input to the simulation; the background was modelled by ran-

[21] In the *XMM* catalogues the detection of the same source in multiple energy bands in the same dataset counts as a single detection. Using this terminology, 2SXPS contains 530,612 detections.

[22] Assuming an absorbed power-law spectrum.

[23] http://cxc.harvard.edu/csc/char.html

Table 6. Catalogs cross-correlated with 2SXPS.

Catalog	Systematic Error[1]	Number of matches[2]	Spurious matches[3]
1SWXRT[4]		35,046	1,427 (4.1%)
1SXPS[5]		98,378	3,223 (3.3%)
2CSC[6]		9,273	602 (6.5%)
2MASS[7]		73,707	43,222 (59%)
2RXS[8]	25″	11,447	1,433 (13%)
3XMM DR8[9]		35,225	3,275 (9.3%)
3XMM Stack[10]		6,938	236 (3.4%)
ALLWISE[11]		156,229	70,543 (45%)
ROSHRI[12]	10″	3,096	365 (12%)
SDSS Quasar Catalog DR14[13]		9,201	75 (0.9%)
SwiftFT[14]		8,985	208 (2.3%)
USNO-B1[15]		128,902	65,539 (51%)
XMM SL2[16]	17″	7,247	2,157 (30%)
XRTGRB[17]		1,188	9 (0.8%)

NOTE—[1] 90% confidence

[2] Number of 1SXPS sources for which there is a counterpart in the external catalog within 3-σ.

[3] The number of 1SXPS sources with a match after the 1SXPS position has been moved by 1–2′; the value in brackets is this number as a percentage of the matches to 1SXPS positions for the same external catalog.

[4] D'Elia et al. (2013); [5] Evans et al. (2014) [6] Evans et al. (2010); [7] Skrutskie et al. (2006) [8] Boller et al. (2016); [9] Rosen et al. (2016), http://xmmssc.irap.omp.eu/Catalogue/3XMM-DR8/3XMM_DR8.html; [10] Traulsen et al. (2019) [11] http://wise2.ipac.caltech.edu/docs/release/allwise/ [12] http://heasarc.gsfc.nasa.gov/W3Browse/rosat/roshri.html [13] Pâris et al. (2018); [14] Puccetti et al. (2011); [15] Monet et al. (2003); [16] Saxton et al. (2008); [17] Taken from http://www.swift.ac.uk/xrt_positions; see Evans et al. (2009);

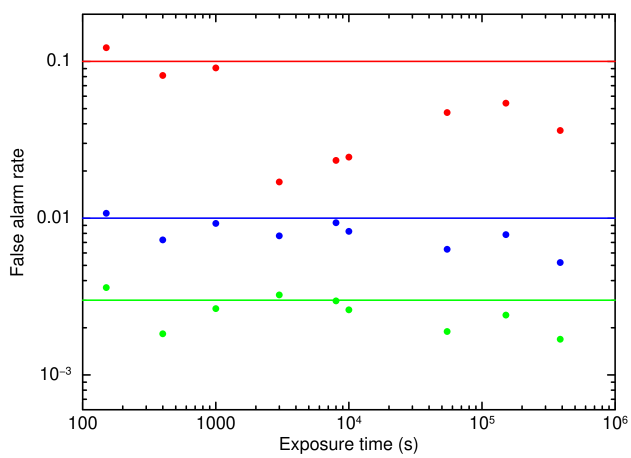

Figure 5. The false positive rate from the simulations as a function of exposure time. The solid lines are at the 0.3%, 1% and 10% levels, and green, blue and red points represent the *Good*, *Good + Reasonable* and complete catalog samples. For some exposure times the false positive rate was never as high as the fiducial value for that flag, so those contamination levels should be treated as conservative.

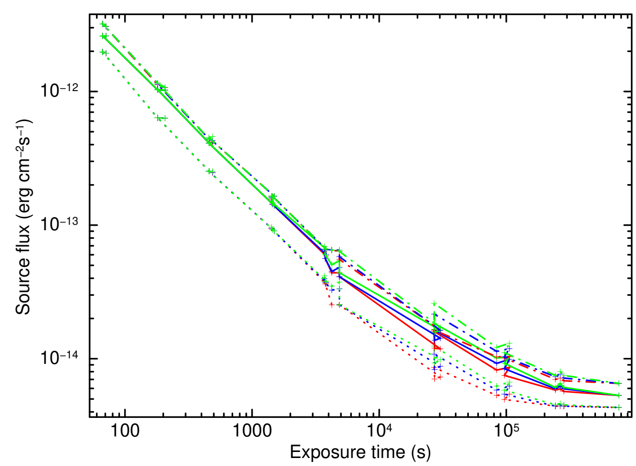

Figure 6. The completeness of the 2SXPS catalog as determined from the simulations. The dotted, solid and dashed lines represent the flux at which 10%, 50%, and 90% of the simulated sources were recovered, as a function of exposure time. The green, blue and red lines represent the *Good*, *Good + Reasonable* and complete catalog samples respectively.

domly drawing the number of counts in each pixel from a Poisson distribution with a mean given by the background map. Sources were added to the image, with their fluxes drawn randomly from the $\log N - \log S$ distribution of extragalactic sources from Mateos et al. (2008). The number of sources per image was also drawn at random from this distribution, although we required a minimum of 10 sources per image to allow us to generate reasonable statistics without requiring a ridiculous number of simulations. We artificially spaced sources to be at least 50 pixels apart, to ensure that the association of detected sources with simulated sources was unambiguous; this may mean that the source completeness in crowded fields is slightly less than from our simulations.

We simulated images with exposures approximately evenly distributed (logarithmically) between 150 s and 1 Ms; for each exposure time[24] we selected three seed datasets from 1SXPS, representing a typical, low and high background level. We then simulated images; the number of simulations performed depended on the exposure time since the shorter images contained fewer sources, but were also quicker to process, details are given in Table 7. Our source detection system was applied to these simulation. Detected sources were either identified with one of the simulated sources (based on position and error), or marked as spurious. We then calibrated the relationship between $L_{\rm src}$, $L_{\rm flat}$, exposure time and detection flag, so as to maximise completeness while obtaining the false positive rates for the different flags as defined in Section 3.5; the resultant thresholds were given in Table 3. Verification of the false positive rate can be seen in Fig. 5. The completeness as a function of exposure time is shown in Fig. 6. This represents a significant increase in sensitivty over paper I: in a 10 ks observation, the flux at which 2SXPS is 50% complete is 3.5 times lower than in 1SXPS. Note that, while we simulated based on three input datasets for each exposure time, and the seed datasets did not have exactly identical exposures, for ease of viewing, we have grouped each set into a single point in these figures.

The reliability of the count rate reconstruction (including effects of the Eddington bias; Eddington 1940), flux estimation using the HRs, and variability estimates were all demonstrated in paper I and we do not repeat that work here.

Table 7. The observations from 1SXPS used as inputs for our simulations.

ObsID	Exposure	BG level[1]	Number of simulations
00032223001	150 s	6.15E-07	20,000
00030051001	150 s	8.46E-07	20,000
00045199001	150 s	1.31E-06	20,000
00031189041	399 s	7.80E-07	20,000
00032433001	399 s	1.45E-06	20,000
00020001001	401 s	5.46E-07	20,000
00047148001	1.0 ks	5.84E-07	6,500
00032200177	1.0 ks	7.36E-07	6,500
00031468029	1.0 ks	1.73E-06	6,500
00035306018	3.0 ks	5.83E-07	3,500
00031142001	3.0 ks	7.47E-07	3,500
00039846003	3.0 ks	1.58E-06	3,500
00037134002	8.0 ks	7.73E-07	1,000
00040508003	8.0 ks	5.93E-07	1,000
00051950063	8.0 ks	1.09E-06	1,000
00037238001	10.0 ks	5.78E-07	1,000
00232683000	10.0 ks	7.80E-07	1,000
00416485007	10.0 ks	1.05E-06	1,000
00302506000	54 ks	5.23E-07	1,000
Stacked im 7508	55 ks	1.11E-06	1,000
00163136014	55 ks	6.86E-07	1,000
Stacked im 7133	150 ks	6.42E-07	1,000
Stacked im 7130	150 ks	7.78E-07	1,000
Stacked im 7616	153 ks	1.35E-06	1,000
Stacked im 5470	360 ks	6.24E-07	1,000
Stacked im 7005	400 ks	7.90E-07	1,000
Stacked im 7032	405 ks	8.30E-07	1,000
Stacked im 7086	1.2 Ms	7.08E-07	1,000

Note—[1] i.e. the mean level in the source-less 1SXPS background map in counts s^{-1} pixel^{-1}.

8. NEXT STEPS

Swift observes a large number of fields every day, and over recent years this observation rate has increased: 2SXPS contains 2.6 times as many observations as 1SXPS, yet only covers 1.5 times as much time (163 months, compared to 107 months). The combination of large sky coverage and good source sensitivity makes the SXPS catalogs a valuable reference to use when identifying possible X-ray transients. That is, when searching for counterparts to gravitational wave triggers many uncatalogued X-ray sources may be found and it is important to know whether they are new transient events or old sources in an area of sky previously uncataloged to XRT-levels of sensitivity.

Due to the delay between an observation being carried out and the data being incorporated in a catalog

[24] Except for 1 Ms, where there was only one 1SXPS field available.

release, rather than waiting some years and then producing 3SXPS, we are instead intending to produce a 'live' *Swift*-XRT Point Source catalog (LSXPS) which will be updated each time a new observation is completed. This will also be a powerful facility for searching for transients or outbursts of known events in real time. This project is in its nascent stages as the moment, however at present we expect that we will do periodic fixed catalog releases (3SXPS, 4SXPS) to provide a reusable and static reference, but these will likely simply be a time-frozen snapshot of LSXPS.

9. CATALOG USAGE

This catalog can be freely used, provided this paper is cited; we also ask users to include the following text in the acknowledgments of any paper using 2SXPS: *This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester.*

ACKNOWLEDGMENTS

PAE, KLP, JPO, and APB acknowledge UKSA support. This research has made use of the XRT Data Analysis Software (XRTDAS) developed under the responsibility of the ASI Space Science Data Center (ASDC), Italy. We made use of the GNU Scientific Libraries (https://www.gnu.org/software/gsl; Galassi et al. 2009); the Minuit2 package provided by CERN (http://project-mathlibs.web.cern.ch/project-mathlibs/sw/Minuit2/html/index.html), the NLOPT fitting library (https://github.com/stevengj/nlopt) and XSPEC. This research has made use of data obtained from the Chandra Source Catalog, provided by the Chandra X-ray Center (CXC) as part of the Chandra Data Archive. 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. This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

We are particularly indebted to Eric Mandel, developer of JS9 (https://js9.si.edu/), for his support during the creation of the 2SXPS website.

REFERENCES

Adrián-Martínez, S., Ageron, M., Albert, A., et al. 2016, JCAP, 2, 062, doi: 10.1088/1475-7516/2016/02/062

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

Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103, doi: 10.1051/0004-6361/201525648

Burrows, D. N., Hill, J. E., Nousek, J. A., et al. 2005, SSRv, 120, 165, doi: 10.1007/s11214-005-5097-2

Calabretta, M. R., & Greisen, E. W. 2002, A&A, 395, 1077, doi: 10.1051/0004-6361:20021327

Cash, W. 1979, ApJ, 228, 939, doi: 10.1086/156922

Churazov, E., Gilfanov, M., Forman, W., & Jones, C. 1996, ApJ, 471, 673, doi: 10.1086/177997

D'Elia, V., Perri, M., Puccetti, S., et al. 2013, A&A, 551, A142, doi: 10.1051/0004-6361/201220863

Ebisawa, K., Bourban, G., Bodaghee, A., Mowlavi, N., & Courvoisier, T. J.-L. 2003, A&A, 411, L59, doi: 10.1051/0004-6361:20031336

Eddington, Sir, A. S. 1940, MNRAS, 100, 354

Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37, doi: 10.1088/0067-0049/189/1/37

Evans, P. A., Beardmore, A. P., Page, K. L., et al. 2009, MNRAS, 397, 1177, doi: 10.1111/j.1365-2966.2009.14913.x

Evans, P. A., Osborne, J. P., Beardmore, A. P., et al. 2014, ApJS, 210, 8, doi: 10.1088/0067-0049/210/1/8

Evans, P. A., Osborne, J. P., Kennea, J. A., et al. 2015, MNRAS, 448, 2210, doi: 10.1093/mnras/stv136

—. 2016a, MNRAS, 455, 1522, doi: 10.1093/mnras/stv2213

Evans, P. A., Kennea, J. A., Palmer, D. M., et al. 2016b, MNRAS, 462, 1591, doi: 10.1093/mnras/stw1746

Evans, P. A., Cenko, S. B., Kennea, J. A., et al. 2017, Science, 358, 1565, doi: 10.1126/science.aap9580

Galassi, M., Davies, J., Theiler, J., et al. 2009, GNU Scientific Library Reference Manual - Third Edition, 3rd edn. (Network Theory Ltd.)

Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, ApJ, 611, 1005, doi: 10.1086/422091

Greisen, E. W., & Calabretta, M. R. 2002, A&A, 395, 1061, doi: 10.1051/0004-6361:20021326

IceCube Collaboration. 2013, Science, 342, 1242856, doi: 10.1126/science.1242856

Kennea, J. A., Coe, M. J., Evans, P. A., Waters, J., & Jasko, R. E. 2018, ApJ, 868, 47, doi: 10.3847/1538-4357/aae839

Kraft, R. P., Burrows, D. N., & Nousek, J. A. 1991, ApJ, 374, 344, doi: 10.1086/170124

Mateos, S., Warwick, R. S., Carrera, F. J., et al. 2008, A&A, 492, 51, doi: 10.1051/0004-6361:200810004

Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, AJ, 125, 984, doi: 10.1086/345888

Moretti, A., Campana, S., Tagliaferri, G., et al. 2004, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5165, SWIFT XRT point spread function measured at the Panter end-to-end tests, ed. K. A. Flanagan & O. H. W. Siegmund, 232–240, doi: 10.1117/12.504857

Moretti, A., Campana, S., Mineo, T., et al. 2005, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5898, UV, X-Ray, and Gamma-Ray Space Instrumentation for Astronomy XIV, ed. O. H. W. Siegmund, 360–368, doi: 10.1117/12.617164

Moretti, A., Pagani, C., Cusumano, G., et al. 2009, A&A, 493, 501, doi: 10.1051/0004-6361:200811197

Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51, doi: 10.1051/0004-6361/201732445

Park, T., Kashyap, V. L., Siemiginowska, A., et al. 2006, ApJ, 652, 610, doi: 10.1086/507406

Pearson, K. 1900, Philosophical Magazine Series 5, 50, 157, doi: 10.1080/14786440009463897

Popp, M., Hartmann, R., Soltau, H., et al. 2000, Nuclear Instruments and Methods in Physics Research A, 439, 567, doi: 10.1016/S0168-9002(99)00912-2

Puccetti, S., Capalbi, M., Giommi, P., et al. 2011, A&A, 528, A122, doi: 10.1051/0004-6361/201015560

Rosen, S. R., Webb, N. A., Watson, M. G., et al. 2016, A&A, 590, A1, doi: 10.1051/0004-6361/201526416

Rowan, T. H. 1990, PhD thesis, University of Texas at Austin, Austin, TX, USA

Saxton, R. D., Read, A. M., Esquej, P., et al. 2008, A&A, 480, 611, doi: 10.1051/0004-6361:20079193

Shaw, A. W., Heinke, C. O., Bahramian, A., et al. 2017, in AAS/High Energy Astrophysics Division, Vol. 16, AAS/High Energy Astrophysics Division #16, 400.01

Singer, L. P., Price, L. R., Farr, B., et al. 2014, Astrophys. J., 795, 105, doi: 10.1088/0004-637X/795/2/105

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163, doi: 10.1086/498708

Smith, R. K., Brickhouse, N. S., Liedahl, D. A., & Raymond, J. C. 2001, ApJL, 556, L91, doi: 10.1086/322992

Traulsen, I., Schwope, A. D., Lamer, G., et al. 2019, A&A, 624, A77, doi: 10.1051/0004-6361/201833938

Vanderbilt, D., & Louie, S. G. 1984, Journal of Computational Physics, 56, 259, doi: 10.1016/0021-9991(84)90095-0

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

Watson, M. G., Schröder, A. C., Fyfe, D., et al. 2009, A&A, 493, 339, doi: 10.1051/0004-6361:200810534

Willingale, R. 2019, in Optics for EUV, X-Ray, and Gamma-Ray Astronomy IX, ed. S. L. O'Dell & G. Pareschi, Vol. 11119, International Society for Optics and Photonics (SPIE), 149 – 165, doi: 10.1117/12.2530682

Willingale, R., Starling, R. L. C., Beardmore, A. P., Tanvir, N. R., & O'Brien, P. T. 2013, MNRAS, 431, 394, doi: 10.1093/mnras/stt175

Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914, doi: 10.1086/317016

Wolter, H. 1952, Annalen der Physik, 445, 94, doi: 10.1002/andp.19524450108

APPENDIX

A. STRAY LIGHT MODELLING

Stray light is a result of the Wolter-I optical design of X-ray telescopes such as the *Swift*-XRT. X-rays originating outside of the nominal field of view undergo a single reflection off the second (hyperbolic) mirror surface, which scatters them onto the detector. The result is a concentric ring pattern on the detector as shown in Fig. 7. Each ring represents reflections off a single mirror shell; the arc-shapes resulting because the X-rays have reflected off a range of azimuthal angles around the mirror, and the ring thickness arising from the extent along the mirror length which can scatter the X-rays onto the CCD.

Willingale (2019) describe in detail how the shape of this pattern can be determined for a given off-axis source position and the geometry of the reflecting surface. Their model was originally produced for the Wide Field Imager instrument on the forthcoming *Athena* satellite, but is applicable to all nested Wolter-I telescopes, such as *Swift*-XRT, and *XMM*. To produce a model for stray light in XRT. we used the equations from Willingale et al. , with details of the XRT mirror from the JET-X design specification[25], which inluded the dimensions, shape and thickness of the mirror shells, baffles and mirror support structure. We then used this model to predict the stray light pattern on the XRT detector in terms of three input parameters: the position angle of the causal source relative to the CCD x axis (θ), the off-axis angle of the source (i.e. the angle between the CCD boresight and the source, ϕ) and a flux normalization (N). The brightness of the rings from was calculated using the X-ray reflectivity of the mirrors, which depends on both photon energy and grazing angle. Note that this model returned the number of counts expected in each CCD pixel as a decimal, i.e. it is not quantised; it therefore served as a model to which the real (quantised and Poisson-distributed) stray light detected could be compared. To perform this comparison, the model image was converted from the CCD detector frame to a sky-coordinate image, using the satellite pointing information in a manner analogous to that used to convert the original event lists into sky images.

An example of the stray light model, converted to sky coordinates, is shown in Fig. 7, along with the actual 0.3–10 keV image. As can readily be seen, the broad features of the data are well reproduced by the model, however there are imperfections: the radius of curvature of the rings is not quite right, and the radial intensity profiles are flatter and wider than the real data. These arise because our model assumes the idealised mirror exactly as per the design, whereas the real mirror has imperfections. The incorrect curvature arises because our model assumes that the XRT mirror shells are perfectly circular in cross section, whereas in reality they are distorted slightly by their connection to the mirror support structure. The radial profile differs from reality because in the idealised model, each mirror shell is perfectly uniform in thickness, and the shells are exactly concentric (i.e. the inter-shell spacing is constant); in the real mirror there are deviations from this idealised scenario which alters radial profile of the rings. A side-effect of the latter problem is that, while the total number of counts in the stray light models was correct, the peak level in the center of the rings was underestimated and so the detection of spurious sources was not adequately reduced. We therefore increased the normalization of the stray light rings by 1.5 compared to that expected from the mirror model (this number reached by trial and improvement). This has the side effect of causing the background to be even more grossly overestimated at the edges of the rings, although in fact this helps to compensate for the curvature errors. Pragmatically, our goal was to suppress the detection of spurious sources resulting from stray light and to flag any detected sources which were likely to be either spurious or at least affected by stray light; the fact that this approach may tend to over-estimate the stray light is prefereable to the alternative.

A.1. *Incorporating stray light into the background model*

When analyzing a dataset containing stray light, the stray light model had to be fitted to the dataset for the reasons discussed in Section 3.2.2. This was a complex process, illustrated in Fig. 8 and described below.

Before any source detection or background modelling was carried out, the snapshots were organized into co-pointed groups. Any snapshot pointed within 80 pixels ($3.1'$) of an earlier snapshot was assigned to the same group as that earlier one[26] Within each group the snapshot with the longest exposure (and so expected to have the best-sampled

[25] The mirrors on XRT were originally fabricated for JET-X.

[26] If a snapshot lay within 80 pixels of multiple disjointed snapshots, it was assigned to whichever group it was closest to.

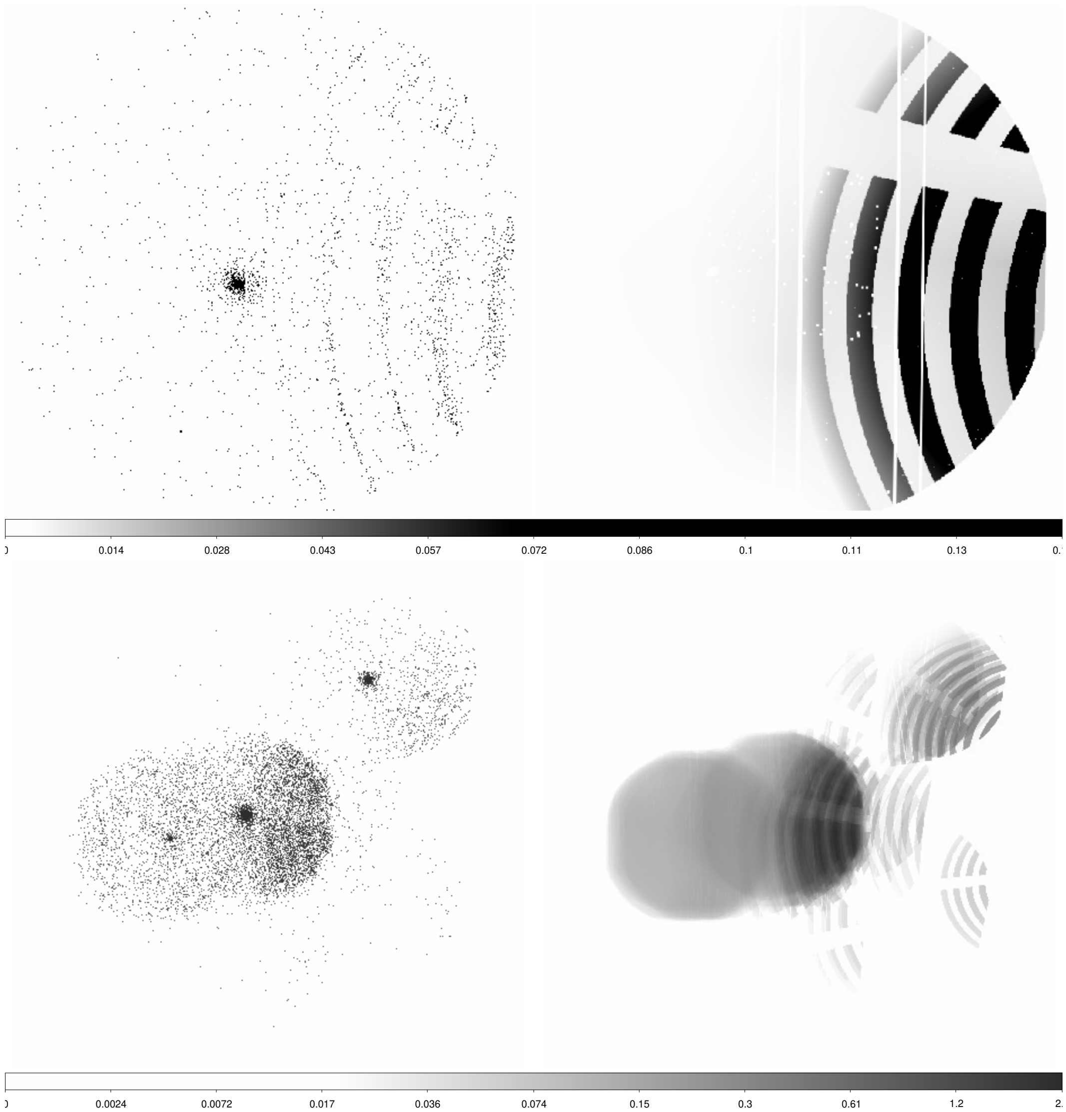

Figure 7. Examples of stray light and its model. *Top*: the data (left) and background map (right) from obsid 00591551000. This dataset contains only a single shapshot of data, so the individual rings are clear. *Bottom*: As the top but for stacked image 14459, which includes the observation from the top panel. A range of obsids are present in this stacked image, many of which suffer stray light contamination from the same source (1SXPS J181228.2-181236). Where obsids have multiple snapshots the effect of the different pointings can be seen as the stray light models overlap, and the shadows caused by the mirror support structure move.

stray light) was identified; these will be referred to hereafter as 'key snapshots'. During the actual background map creation (below), the full fit and test of whether stray light was needed was carried out only for the key snapshots; for the other snapshots, only the normalization was fitted: the position was fixed. This was primarily for reasons of computational efficiency: the fitting process was CPU intensive and slow, thus by reducing the number of snapshots for which the full fit was needed, the overall runtime could be significantly reduced.

Not all steps in the fitting process (Fig. 8) were carried out each time the background map was constructed, as indicated by the decision forking. Here we describe the essential algorithm, with the deviations from it explained afterwards. Note that this presupposes that (a) potential source(s) of stray light had been identified as describe above (Section 3.2.1); if not, none of the stray-light-specific steps described here were carried out. Some steps are indicated with lower-case Roman numerals below and in Fig. 8, for ease of reference later on.

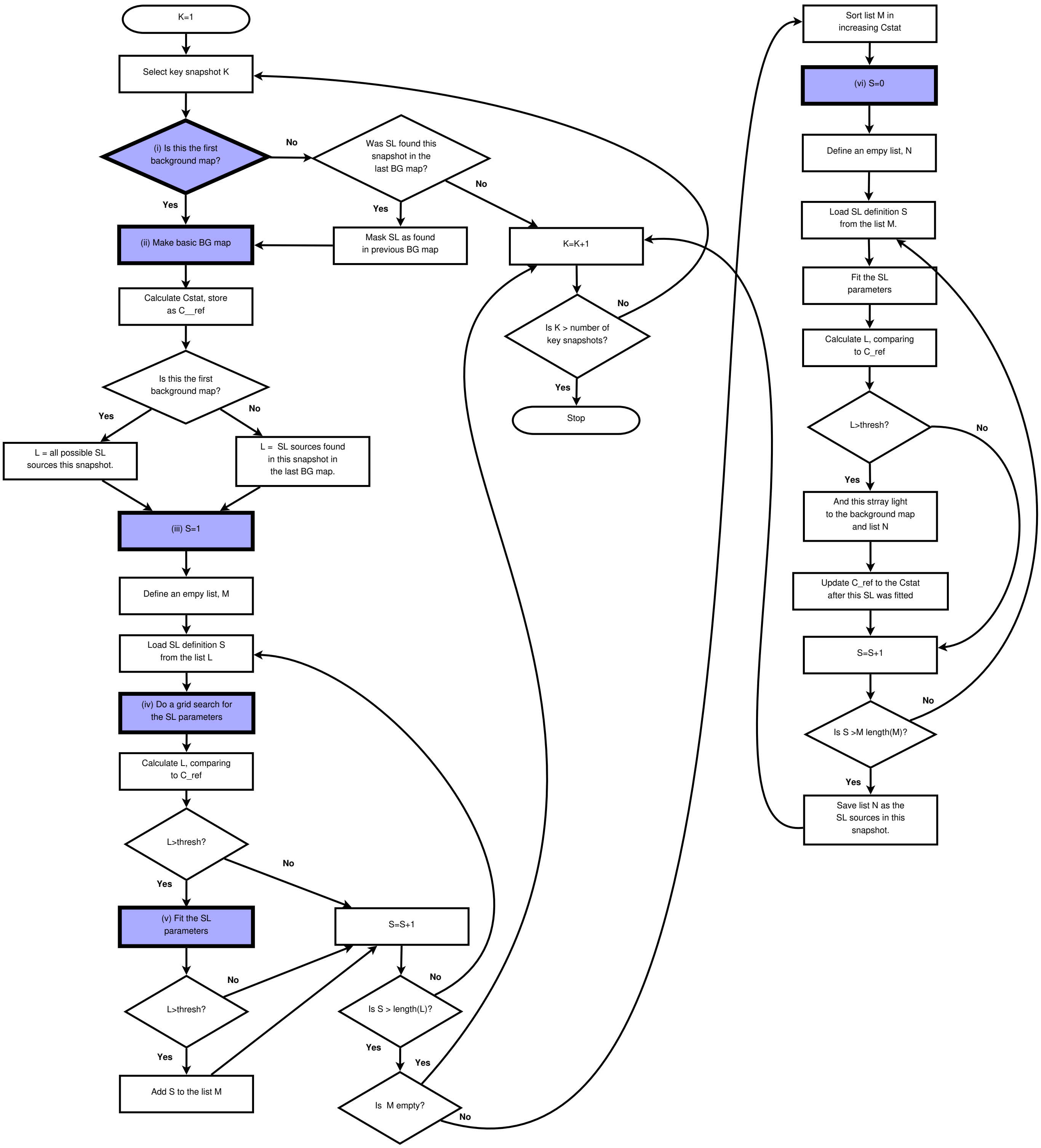

Figure 8. Flow-chart depicting the algorithm used for searching for stray light, and including a model of it in the background map. Blue boxes with heavy borders mark the reference points numbered (lower-case Roman numerals) in the text.

The construction of the background map, as in paper I, consisted of iterating over all snapshots, creating a map per snapshot, and then summing them. For datasets with possible stray light, the key snapshots were processed first.

(i) For each key snapshot, as well as masking out any sources already detected, the regions of the CCD covered by stray light, as modelled last time the background map was created, were also masked out. **(ii)** The 'basic map' (i.e. that created by the mask/rebin/interpolate approach) was then created. The C-stat was calculated (Equation 1) comparing this background map with the snapshot image data, this value was recorded as $\mathcal{C}_{\rm ref}$.

(iii) The possible stray light sources were then considered independently. The first time the background map was created, the positions of the stray-light sources were converted from (α, δ) to (θ, ϕ), these being the parameters to be fitted and stored internally. θ, the position angle from the CCD x axis to the source was allowed to vary by $\pm 5°$; ϕ, the off-axis angular distance to the source was given a range $\pm 10'$. The normalization was fitted in log space, and allowed to vary by ± 3 dex from the initial estimated value (determined from the cataloged flux of the source). **(iv)** Initially a grid search was performed to determine the best starting point for a fit. The three parameters were stepped over their ranges in 5 steps, $\mathcal{C}$ calculated at each point, at the best parameters and $\mathcal{C}$ noted. A likelihood test (Equation 2) was carried out comparing this best $\mathcal{C}$ with $\mathcal{C}_{\rm ref}$ determined in step (ii) and unless L was at least 15, stray light was deemed not to be present from this source and it was ignored. **(v)** For cases where $L \geq 15$, a fit was performed, using the best parameters from the grid search as the starting point, but retaining the parameter limits from step (iii). Fitting was carried out using the NLOPT library[27] and the NL_SBPLX algorithm, which is based on the 'subplex' algorithm of Rowan (1990). L was calculated comparing $\mathcal{C}$ from the best fit with $\mathcal{C}_{\rm ref}$, and if $L \geq 30$ the stray light source was saved as a *possible* contributor to stray light in this snapshot.

(vi) Once steps (iii)–(v) had been performed for each possible stray light source, any which passed the likelihood test were sorted into decreasing order of fit quality (i.e. increasing order of $\mathcal{C}$). These were again fitted as in step (v), except that this time the likelihood threshold was increased to 32. If a stray light source passed this threshold, it was deemed to be present in the data. The best-fitting model of the stray light was immediately added to the background map used to fit the next possible stray light source, and $\mathcal{C}_{\rm ref}$ was set to the $\mathcal{C}$ value found from the fit. Thus, once a stray light source had been identified a subsequent source could only also be added to a key snapshot if it was still deemed significant given the presence of the source(s) already identified. This was necessary because even a badly-fitting or unnecessary stray light model gave a significant improvement to $\mathcal{C}$ when the true source of the stray light is not included in the model.

(vii) For any sources of stray light which passed all of the above steps, the best fitting position parameters (θ, ϕ) were compared to the best-fitting values from the last time a background map was created. If the new position represented a shift in (α, δ) of $30''$ or more (or if this was the first background map, so no previously fitted stray light positions existed), it is likely that the stray light was incorrectly masked out during step (iii). So the entire process [steps (iii)–(vi)] was repeated, using these new positions as the starting point, and for masking. Note that all stray light sources that had been tested during steps (iii)–(vi) were included again in this pass, including those where L had been found to be below threshold; this is in case the improved masking changed the L values.

After the above process had been carried out for each key snapshot, and the resultant stray light sources and their parameters stored, the remaining snapshots were processed. For these, the stray light definitions were taken from the relevant key snapshot, masked out in the creation of the basic map, and then fitted using the same library as above [step (v)], but this time only the normalization was free to vary and only by ± 2 dex. No likelihood tests were performed: all stray light sources accepted for the key snapshot were modelled in each snapshot within that group.

The above algorithm describes the overall approach followed each time a background map was created during the source detection process, however there were deviations from this approach. The list of possible stray-light sources used from step (iii) onwards was not constant. In the first background map, all possible stray light sources identified when the data were being prepared were considered; in subsequent maps, only sources which passed the likelihood tests in steps (iv)–(vi) were used in step (iii) of the next background map. During the very first background map creation, the thresholds used in steps (iv)–(vi) were all set to one initially; i.e. any possible stray light source that made even a marginal improvement to the background model was initially retained – this was because at this point no stray light had been masked in creation of the basic map, which could significantly reduce the L values returned. However, once step (vii) was reached even on the first background map, the L thresholds were restored to those described above.

[27] https://github.com/stevengj/nlopt

During source detection, once all of the high S/N sources had been detected and the S/N threshold reduced (i.e. phase two, middle column of Fig. 2, had begun) the positions and number of stray light sources was fixed completely; hereafter the key snapshots were handled like the normal snapshots, i.e. only the normalization was able to be refitted.

The full stray-light fitting process described above was only carried out on the total (0.3–10 keV) band image, and only for individual observations. In the former case, this is the image with the most events and so in which the stray light could be best modelled. Since the other images are subsets of the total-band image, it is nonsensical to independently fit the stray light; instead the positions of the stray light sources (per snapshot) from the total image were supplied to the source detection code for the soft, medium and hard bands and all snapshots were treated as non-key snapshots, i.e. only the stray-light normalization was fitted, and no likelihood tests were carried out.

For stacked images it was not necessary to carry out the full stray light fitting, since the background mapping works on individual snapshots, regardless of what type of dataset is being analysed[28] Therefore, for stacked images only the stray light normalization was fitted, and only the stray light sources necessary for the component observations were used, with no likelihood tests performed.

A shortcoming of our approach is that it will not pick up stray light too faint to make a significant impact on an individual snapshot, but which is visible – and produces spurious detections – in the full image for the dataset. In reality, this situation is rare, and was caught during the visual screening phase (Section 3.6). The only alternative would be to simultaneously fit all snapshots, which is not practical.

B. PSF CALIBRATION

Calibrating the PSF wings (i.e. the regions more than ~ 30 pixels from the source) is challenging, since these contain only a small fraction of the source flux. Bright sources cannot be used for this calibration as their PSF shape is distorted by pile up. Instead one must identify modest-brightness sources with long exposures. This is also problematic for *Swift* because it has a low-Earth orbit, therefore long exposures can only be achieved by comibining data from multiple spacecraft pointings. The star trackers on-board *Swift* give astrometric accuracy of $3.5''$ (90% confidence), compared to a pixel size of $2.357''$, thus when combining the data one must account for the fact that the source position can move slightly between observations, which will both broaden the PSF and change its shape. So, we require sources bright enough for a sub-pixel localisation to be performed for each snapshot.

We therefore selected sources in the 1SXPS catalog with total-band count rates in the range 0.3–0.6 s^{-1}, a minimum of five separate observations in the catalog, and a Galactic absorption column below 3×10^{21} cm^{-2}. The latter criterion was to reduce the risk of high foreground dust contamination which can broaden the PSF by scattering. For each source we identified the stacked image in 1SPXS it was in, and used the data from that to model the PSF. The data were split into snapshots, and we performed a source centroid on each snapshot individually, rejecting any snapshots where the 1-σ position error was above 0.5 pixels.

We simultaneously fitted the same model to all snapshots individually, where the source position for each snapshot was taken from the centroiding performed above as any form of shifting and adding the individual artificially broaden the PSF. We found that the fits tended to be prone to local minima and therefore fitted the PSF profile using the simulated annealing approach of Vanderbilt & Louie (1984). The fitted model was that given in Equation 3. The fits to some of the sources gave parameters significantly discrepent from those obtained from the majority of sources, likely indicating either a failure to find the true minimum, or possibly some issue with the data (such as a dust-scattering halo, or contamination from an unresolved nearby source). We tried refitting with a slower 'cooling rate' for the simulated anneal, and if the result was still strongly discrepent, we excluded the source from the analysis. This left us with 25 sources with similar PSF fit results to each other.

For each parameter in the PSF, we combined the best-fit values and uncertainties from these fits to produce a probability distribution function which, due to the central limit theorem, we expect to be Gaussian in nature. We then modelled this with a Gaussian function, to produce the best-fitting parameters used for 2SXPS, which were given in Table 2. Unlike the current CALDB parameters, we find that a Gaussian component is necessary; as can be seen in Fig. 9, its inclusion improves the modelling of the PSF core, particularly at 8–15 pixels, which in turn allows the King component to broaden, better reproducing the wings.

[28] In principle a stacked image could yield more sources with S/N> 10 which could have a small effect on the stray light position. Such effect is small however, and the stray light fitting is so computationally demanding that it is not sensible to run it independently on the stacked image, for a negligible improvement.

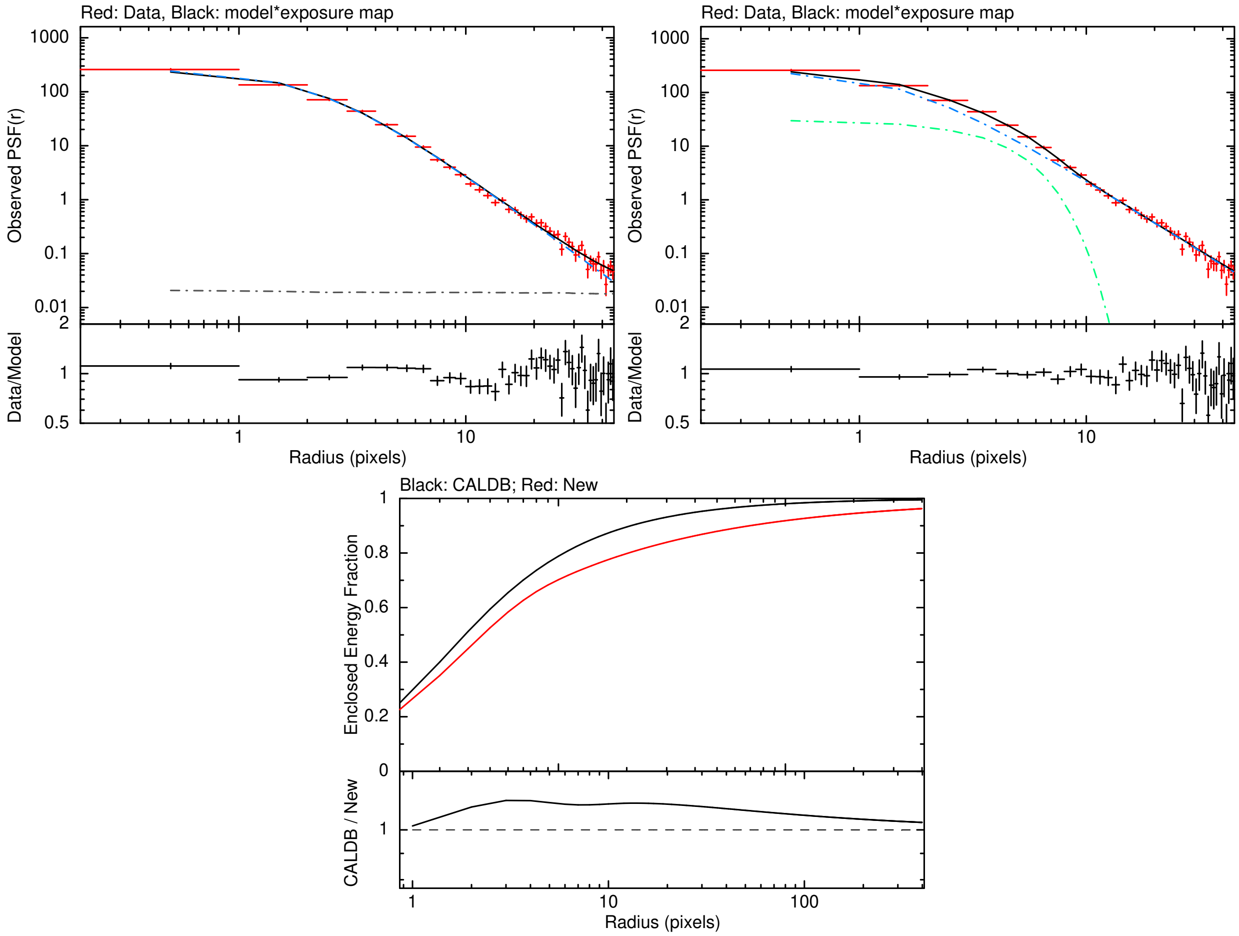

Figure 9. The CALDB (top-left) PSF model and our new PSF model (top-right) fitted to the same dataset. The blue, green and black dashed lines show the Gaussian and King components and fitted background. The improvement in residuals around 8–15 pixels can clearly be seen. With the CALDB PSF, the counts beyond ~ 30 pixels are all background counts; the new PSF has broader wings which reproduce the events out to large radii (the background is below the y-axis in this plot). *bottom*: The enclosed energy fraction of the CALDB (black) and new (red) PSF models, with the ratio of the two in the bottom panel.

The PSF CALDB file allows for the PSF parameters to vary with energy, off-axis angle and the product of these properties. Such variation, especially with energy, is expected physically, and was measured in the ground calibration (Moretti et al. 2004). All of the sources we analysed were close to on-axis, however we split the data into different energy bands and repeated the fitting process. Unfortunately, due to the lower number of counts per band the uncertainties on the parameters we derived were large. No evidence was seen for energy-dependence but with no strong constraints. We elected therefore to treat the PSF as being independent of energy for the purposes of our catalogue construction. Since most modest-brightness sources are the target of their observation and are therefore on-axis, we lack the data to properly calibrate the off-axis angle dependence, so this was also ignored.

C. CATALOG TABLES

There are four 2SXPS tables for download. The contents of these files are given in the following tables. The files are available in 3 formats: as a comma separated values (csv) file, a FITS file, and as an SQL dump (MYSQL/MARIADB format).

The primary catalog product is the '2SXPS_Sources' file which contains details of the unique sources, as described in Table 8. '2SXPS_Datasets' (Table 9) describes the individual datasets; '2SXPS_Detections' (Table 10) gives details of all of the individual detections and '2SXPS_xCorr' (Table 11) lists all the external catalog matches.

Table 8. Contents of the main catalog table ('sources'), containing an entry per unique source detected in 2SXPS.

Field	Units	Description	Has errors?[1]
		Name and position	
2SXPS_ID		Numerical unique source identifier within 2SXPS	
IAUName		IAU-format name, 2SXPS JHHMMSS[+-]ddmmsss	
RA	Deg	Right Ascension (J2000)	
Decl	Deg	Declination (J2000)	
Err90	arcsec	Position uncertainty, 90% confidence, radial, assumed to be Rayleigh-distributed	
AstromType		Provenance of source astrometry. 0=Swift star tracker, 1=XRT+2MASS astrometry	
l	Deg	Galactic longitude	
b	Deg	Galactic latitude	
MeanOffAxisAngle	arcmin	The mean angular distance of the source from the XRT boresight in all observations in which the source was detected	
NearestNeighbour	arcsec	The distance to the closest 2SXPS source to this one	
NearestOKNeighbour	arcsec	The distance to the closest 2SXPS source to this one which is ranked Good or Reasonable and has no other DetFlag bits set	
		Exposure details	
Exposure	s	The total exposure at the source location in the catalog	
FirstObsDate	UTC	The time of the start of the first observation in 2SXPS which covered the source location	
LastObsDate	UTC	The time of the end of the last observation in 2SXPS which covered the source location	
FirstObsMET	MET[2]	The time of the start of the first observation in 2SXPS which covered the source location	
LastObsMET	MET	The time of the end of the last observation in 2SXPS which covered the source location	
FirstDetDate	UTC	The date & time of the start of the first observation in 2SXPS in which the source count rate is inconsisent with 0 at the 3-σ level	
LastDetDate	UTC	The date & time of the end of the last observation in 2SXPS in which the source count rate is inconsisent with 0 at the 3-σ level	
FirstDetMET	MET	The time of the start of the first observation in 2SXPS in which the source count rate is inconsisent with 0 at the 3-σ level	
LastDetMET	MET	The time of the end of the last observation in 2SXPS in which the source count rate is inconsisent with 0 at the 3-σ level	

Table 8 *continued*

Table 8 *(continued)*

Field	Units	Description	Has errors?[1]
FirstBlindDetDate	UTC	The UTC date & time of the start of the first observation in 2SXPS in which the source is detected in the blind search	
LastBlindDetDate	UTC	The UTC date & time of the end of the last observation in 2SXPS in which the source is detected in the blind search	
FirstBlindDetMET	MET	The time of the start of the first observation in 2SXPS in which the source is detected in the blind search,	
LastBlindDetMET	MET	The time of the end of the last observation in 2SXPS in which the source is detected in the blind search.	
NumObs		The number of observations covering this source's position	
NumBlindDetObs		The number of observations in which this source was found in a blind search.	
NumDetObs		The number of observations in which this source is detected.	
BestDetectionID		The ID of the detection from which the position and error were taken (cf the detections table).	
NonBlindDet_band0	[Bool]	Whether the count rate in the total band is inconsistent with 0 at the 3-σ level (0 for no, 1 for yes).	
NonBlindDet_band1	[Bool]	Whether the count rate in the soft band is inconsistent with 0 at the 3-σ level (0 for no, 1 for yes).	
NonBlindDet_band2	[Bool]	Whether the count rate in the medium band is inconsistent with 0 at the 3-σ level (0 for no, 1 for yes).	
NonBlindDet_band3	[Bool]	Whether the count rate in the hard band is inconsistent with 0 at the 3-σ level (0 for no, 1 for yes).	
		Flag details	
DetFlag		The overall source detection flag	
FieldFlag		The best field flag from all detections of this source	
DetFlag_band0		The overall detection flag the total band	
DetFlag_band1		The overall detection flag the soft band	
DetFlag_band2		The overall detection flag in the medium band	
DetFlag_band3		The overall detection flag in the hard band	
OpticalLoadingWarning	Mag	The worst optical loading warning from all detections	
StrayLightWarning	[Bool]	Whether any detection of this source occurred within fitted stray light rings.	
NearBrightSourceWarning	[Bool][4]	Whether any detection of this source occurred within the PSF wings of a bright object.	
IsPotentialAlias		Whether the source is likely aliased with other sources	
PotentialAliasList		The 2SXPS_IDs of any sources which may be aliases of this	
		Count rate and variability information	
Rate_band0	s^{-1}	The mean count rate in the total band	Yes
HR1		The aggregate HR1 hardness ratio of the source	Yes
HR2		The aggregate HR2 hardness ratio of the source	Yes

Table 8 *continued*

Table 8 *(continued)*

Field	Units	Description	Has errors?[1]
Rate_band1	s^{-1}	The mean count rate in the soft band	Yes
Rate_band2	s^{-1}	The mean count rate in the medium band	Yes
Rate_band3	s^{-1}	The mean count rate in the hard band	Yes
Counts_band0		The total number of counts in the source region in the total band	
Counts_band1		The total number of counts in the source region in the soft band	
Counts_band2		The total number of counts in the source region in the medium band	
Counts_band3		The total number of counts in the source region in the hard band	
BgCounts_band0		The total number of background counts expected in the source region in the total band	
BgCounts_band1		The total number of background counts expected in the source region in the soft band	
BgCounts_band2		The total number of background counts expected in the source region in the medium band	
BgCounts_band3		The total number of background counts expected in the source region in the hard band	
RateCF_band0		The PSF correction factor in the total band	
RateCF_band1		The PSF correction factor in the soft band	
RateCF_band2		The PSF correction factor in the medium band	
RateCF_band3		The PSF correction factor in the hard band	
UL_band0	s^{-1}	The 3-σ upper limit on the count rate in the total band	
UL_band1	s^{-1}	The 3-σ upper limit on the count rate in the soft band	
UL_band2	s^{-1}	The 3-σ upper limit on the count rate in the medium band	
UL_band3	s^{-1}	The 3-σ upper limit on the count rate in the hard band	
PeakRate_band0[5]	s^{-1}	The peak count rate in the total band	Yes
PeakRate_band1[5]	s^{-1}	The peak count rate in the soft band	Yes
PeakRate_band2[5]	s^{-1}	The peak count rate in the medium band	Yes
PeakRate_band3[5]	s^{-1}	The peak count rate in the hard band	Yes
PvarPchiSnapshot_band0		The probability that the source count rate in the total band does not vary between snapshots	
PvarPchiSnapshot_band1		The probability that the source count rate in the soft band does not vary between snapshots	
PvarPchiSnapshot_band2		The probability that the source count rate in the medium band does not vary between snapshots	
PvarPchiSnapshot_band3		The probability that the source count rate in the hard band does not vary between snapshots	
PvarPchiSnapshot_HR1		The probability that the source HR1 hardness ratio does not vary between snapshots	
PvarPchiSnapshot_HR2		The probability that the source HR2 hardness ratio does no t vary between snapshots	
PvarPchiObsID_band0		The probability that the source count rate in the total band	

Table 8 *continued*

Table 8 *(continued)*

Field	Units	Description	Has errors?[1]
		does not vary between observations	
PvarPchiObsID_band1		The probability that the source count rate in the soft band does not vary between observations	
PvarPchiObsID_band2		The probability that the source count rate in the medium band does not vary between observations	
PvarPchiObsID_band3		The probability that the source count rate in the hard band does not vary between observations	
PvarPchiObsID_HR1		The probability that the source HR1 hardness ratio does not vary between observations	
PvarPchiObsID_HR2		The probability that the source HR2 hardness ratio does not vary between observations	
		Flux and spectral information	
GalacticNH	cm^{-2}	The Galactic absorption column in the direction of the source, from Willingale et al (2013)	
WhichPow		Which method of determining the spectral properties assuming a power-law was used	
WhichAPEC		Which method of determining the spectral properties assuming an APEC was used	
PowECFO	$erg\ cm^{-2}ct^{-1}$	The observed flux ECF[3], assuming a power-law spectrum.	
PowECFU	$erg\ cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming a power-law spectrum.	
PowFlux	$erg\ cm^{-2}s^{-1}$	The mean total observed flux assuming a power-law spectrum.	Yes
PowUnabsFlux	$erg\ cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming a power-law spectrum.	Yes
APECECFO	$erg\ cm^{-2}ct^{-1}$	The observed flux ECF, assuming an APEC spectrum.	
APECECFU	$erg\ cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming an APEC spectrum.	
APECFlux	$erg\ cm^{-2}s^{-1}$	The mean total observed flux assuming an APEC spectrum.	Yes
APECUnabsFlux	$erg\ cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming an APEC spectrum.	Yes
PowPeakFlux	$erg\ cm^{-2}s^{-1}$	The peak total observed flux assuming a power-law spectrum.	Yes
PowPeakUnabsFlux	$erg\ cm^{-2}s^{-1}$	The peak total unabsorbed flux assuming a power-law spectrum.	Yes
APECPeakFlux	$erg\ cm^{-2}s^{-1}$	The peak total observed flux assuming an APEC spectrum.	Yes
APECPeakUnabsFlux	$erg\ cm^{-2}s^{-1}$	The peak total unabsorbed flux assuming an APEC spectrum.	Yes
FixedPowECFO	$erg\ cm^{-2}ct^{-1}$	The observed flux ECF, assuming the canned power-law spectrum.	
FixedPowECFU	$erg\ cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming the canned power-law spectrum.	
FixedPowFlux	$erg\ cm^{-2}s^{-1}$	The mean total observed flux assuming the canned power-law spectrum.	Yes
FixedPowUnabsFlux	$erg\ cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming the canned power-law spectrum.	Yes
FixedAPECECFO	$erg\ cm^{-2}ct^{-1}$	The observed flux ECF, assuming the canned APEC spectrum.	
FixedAPECECFU	$erg\ cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming the canned APEC spectrum.	
FixedAPECFlux	$erg\ cm^{-2}s^{-1}$	The mean total observed flux assuming the canned APEC spectrum.	Yes
FixedAPECUnabsFlux	$erg\ cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming the canned APEC spectrum.	Yes
InterpPowECFO	$erg\ cm^{-2}ct^{-1}$	The observed flux ECF, assuming the power-law spectrum interpolated from the HRs.	

Table 8 *continued*

Table 8 *(continued)*

Field	Units	Description	Has errors?[1]
InterpPowECFU	erg $cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming the power-law spectrum interpolated from the HRs.	
InterpPowNH	cm^{-2}	The hydrogen column density inferred assuming the power-law spectrum interpolated from the HRs.	Yes
InterpPowGamma		The spectral photon index inferred assuming the power-law spectrum interpolated from the HRs.	Yes
InterpPowFlux	erg $cm^{-2}s^{-1}$	The mean total observed flux assuming the power-law spectrum interpolated from the HRs.	Yes
InterpPowUnabsFlux	erg $cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming the power-law spectrum interpolated from the HRs.	Yes
InterpAPECECFO	erg $cm^{-2}ct^{-1}$	The observed flux ECF, assuming the APEC spectrum interpolated from the HRs.	
InterpAPECECFU	erg $cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming the APEC spectrum interpolated from the HRs.	
InterpAPECNH	cm^{-2}	The hydrogen column density inferred assuming the APEC spectrum interpolated from the HRs.	Yes
InterpAPECkT	keV	The temperature inferred assuming the APEC spectrum interpolated from the HRs.	Yes
InterpAPECFlux	erg $cm^{-2}s^{-1}$	The mean total observed flux assuming the APEC spectrum interpolated from the HRs.	Yes
InterpAPECUnabsFlux	erg $cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming the APEC spectrum interpolated from the HRs.	Yes
P_pow		The probability that the HR values of this source could be obtained if the true spectrum is an absorbed power-law	
P_APEC		The probability that the HR values of this source could be obtained if the true spectrum is an APEC.	
FittedPowECFO	erg $cm^{-2}ct^{-1}$	The observed flux ECF, assuming the power-law spectral model fitted to a custom-built spectrum.	
FittedPowECFU	erg $cm^{-2}ct^{-1}$	The unabsorbed flux ECF, assuming the power-law spectral model fitted to a custom-built spectrum.	
FittedPowNH	cm^{-2}	The hydrogen column density inferred assuming the power-law spectral model fitted to a custom-built spectrum.	Yes
FittedPowGamma		The spectral photon index inferred assuming the power-law spectral model fitted to a custom-built spectrum.	Yes
FittedPowFlux	erg $cm^{-2}s^{-1}$	The mean total observed flux assuming the power-law spectral model fitted to a custom-built spectrum.	Yes
FittedPowUnabsFlux	erg $cm^{-2}s^{-1}$	The mean total unabsorbed flux assuming the power-law spectral model fitted to a custom-built spectrum.	Yes
FittedPowCstat		The C-statistic from the power-law spectral fit to the custom-built spectrum.	

Table 8 *continued*

Table 8 *(continued)*

Field	Units	Description	Has errors?[1]
FittedPowDOF		The number of degrees of freedom in the power-law spectral fit to the custom-built spectrum.	
FittedPowReducedChi2		The Churazov-weighted reduced χ^2 from the power-law spectral fit to the custom-built spectrum.	
FittedAPECECFO	erg cm^{-2}ct^{-1}	The observed flux ECF, assuming the APEC spectral model fitted to a custom-built spectrum.	
FittedAPECECFU	erg cm^{-2}ct^{-1}	The unabsorbed flux ECF, assuming the APEC spectral model fitted to a custom-built spectrum.	
FittedAPECNH	cm^{-2}	The hydrogen column density inferred assuming the APEC spectral model fitted to a custom-built spectrum.	Yes
FittedAPECkT	keV	The temperature inferred assuming the APEC spectral model fitted to a custom-built spectrum.	Yes
FittedAPECFlux	erg cm^{-2}s^{-1}	The mean total observed flux assuming the APEC spectral model fitted to a custom-built spectrum.	Yes
FittedAPECUnabsFlux	erg cm^{-2}s^{-1}	The mean total unabsorbed flux assuming the APEC spectral model fitted to a custom-built spectrum.	Yes
FittedAPECCstat		The C-statistic from the APEC spectral fit to the custom-built spectrum.	
FittedAPECDOF		The number of degrees of freedom in the APEC spectral fit to the custom-built spectrum.	
FittedAPECReducedChi2		The Churazov-weighted reduced χ^2 from the APEC spectral fit to the custom-built spectrum.	
HasSpec		Whether a custom-built spectrum was created for this source.	
Cross-correlation information			
NumExternalMatches		The number of external sources found to agree spatially with this one at the 3-σ level.	
NumExternalMatches_slim		The number of external sources found to agree spatially with this one at the 3-σ level, excluding 2MASS, USNO-B1 and ALLWISE matches.	
MatchInROSHRI	[Bool]	Whether the source has a match in ROSAT HRI	
MatchIn2RXS	[Bool]	Whether the source has a match in 2RXS	
MatchIn3XMMDR8	[Bool]	Whether the source has a match in 3XMM-DR8	
MatchIn3XMM_Stack	[Bool]	Whether the source has a match in 3XMM-DR7s	
MatchInXMMSL2	[Bool]	Whether the source has a match in XMMSL2	
MatchInSwiftFT	[Bool]	Whether the source has a match in SwiftFT	
MatchIn1SWXRT	[Bool]	Whether the source has a match in 1SWXRT	
MatchInXRTGRB	[Bool]	Whether the source has a match in the XRT GRB afterglows.	
MatchInSDSSQSO	[Bool]	Whether the source has a match in SDSS QSO DR14	
MatchIn2MASS	[Bool]	Whether the source has a match in 2MASS	
MatchInUSNOB1	[Bool]	Whether the source has a match in USNO-B1	

Table 8 *continued*

Table 8 *(continued)*

Field	Units	Description	Has errors?[1]
MatchIn2CSC	[Bool]	Whether the source has a match in 2CSC	
MatchIn1SXPS	[Bool]	Whether the source has a match in 1SXPS	
MatchInALLWISE	[Bool]	Whether the source has a match in ALLWISE	

Note— Boolean columns (marked as '[Bool]' above) have a value of 0 for false and 1 for true.
[1] This is 'no' unless stated. For a field with errors, there are two error fields, *fieldname*_pos and *fieldname*_neg.
[2] MET = Swift Mission Elapsed Time = Seconds since 2001 Jan 01 00:00:00 (TT)
[3] ECF = Energy Conversion Factor, i.e. the conversion from observed 0.3–10 keV counts to 0.3–10 keV flux; ECFs are provided to convert to observed and unabsorbed flux.
[4] NearBrightSourceWarning can have a value of 2, as discussed in Section 3.5.
[5] The peak rate was defined in Section 4.

Table 9. Contents of the 'Datasets' catalog table, containing an entry per dataset in the catalog

Field	Units	Description
ObsID[1]		*Swift* obsID of the dataset
FieldFlag		The warning flag associated with this dataset
RA	deg	The Right Ascension (J2000) of the dataset center
Decl	deg	The declination (J2000) of the dataset center
l	deg	Galactic longitude of the dataset center
b	deg	Galactic latitude of the dataset center
ImageSize	pix	The side length of the dataset image in XRT pixels
ExposureUsed	s	The post-filtering exposure in the dataset
OriginalExposure	s	The original exposure in the dataset
StartTime_MET	MET	The start time of the dataset
StopTime_MET	MET	The end time of the dataset
MidTime_MET	MET	The mid-time of the dataset
MidTime_TDB	TDB	The mid-time of the dataset
MidTime_MJD	MJD	The mid-time of the dataset
StartTime_UTC	UTC	The start time of the dataset
StopTime_UTC	UTC	The end time of the dataset
FieldBG_band0	ct s^{-1} pix^{-1}	The mean background level in the total band
FieldBG_band1	ct s^{-1} pix^{-1}	The mean background level in the soft band
FieldBG_band2	ct s^{-1} pix^{-1}	The mean background level in the medium band
FieldBG_band3	ct s^{-1} pix^{-1}	The mean background level in the hard band
NumSrc_band0		The number of sources detected in this dataset in the total band
NumOK_band0		The number of *Good* or *Reasonable* sources detected in this dataset in the total band
MedianDist_band0	arcsec	The median distance between sources detected in

Table 9 *continued*

Table 9 *(continued)*

Field	Units	Description
		this dataset in the total band
NumSrc_band1		The number of sources detected in this dataset in the soft band
NumOK_band1		The number of good or reasonable sources detected in this dataset in the soft band
MedianDist_band1	arcsec	The median distance between sources detected in this dataset in the soft band
NumSrc_band2		The number of sources detected in this dataset in the medium band
NumOK_band2		The number of good or reasonable sources detected in this dataset in the medium band
MedianDist_band2	arcsec	The median distance between sources detected in this dataset in the medium band
NumSrc_band3		The number of sources detected in this dataset in the hard band
NumOK_band3		The number of good or reasonable sources detected in this dataset in the hard band
MedianDist_band3	arcsec	The median distance between sources detected in this dataset in the hard band
NumberOfSnapshots		The number of snapshots contributing to this dataset
AstromError	arcsec	The 90% confidence radial uncertainty on the XRT-2MASS astrometric solution
CRVAL1_corr		The CRVAL1 WCS reference value for the dataset derived from the XRT-2MASS astrometric solution
CRVAL2_corr		The CRVAL2 WCS reference value for the dataset derived from the XRT-2MASS astrometric solution
CROTA2_corr		The CROTA1 WCS reference value for the dataset derived from the XRT-2MASS astrometric solution

Note—[1] Values $> 10^{10}$ refer to stacked images.

Table 10. Contents of the 'Detections' catalog table, containing an entry per detection in the catalogue

Field	Units	Description	Has errors?
DetectionID		A unique identifier for this detection	
2SXPS_ID		The 2SXPS sourceID with which this detection is associated	
SourceNo		The identifier of this source within this obsid and band	
Band		The energy band in which this detection occurred	

Table 10 *continued*

Table 10 *(continued)*

Field	Units	Description	Has errors?
ObsID		The identifier of the observation or stacked image in which this detection occurred.	
CorrectedExposure	s	The exposure time at the position of the source in this obsID	
ExposureFraction	s	The fractional exposure at the position of this source, i.e. the exposure divided by the nominal exposure for the field	
OffaxisAngle	arcmin	The angular distance of the source from the XRT boresight	
RA	deg	Right Ascension (J2000)	Yes
Decl	deg	Decliniation (J2000)	Yes
Err90	arcsec	Position uncertainty, 90% confidence, radius	
RA_corrected	deg	Right Ascension (J2000) using XRT-2MASS astrometry	
Decl_corrected	deg	Declination (J2000) using XRT-2MASS astrometry	
Err90_corrected	arcsec	Uncertainty on the position, 90% confidence radius	
l	deg	Galactic longitude	
b	deg	Galactic latitude	
l_corrected	deg	Galactic longitude using XRT-2MASS astrometry	
b_corrected	deg	Galactic latitude using XRT-2MASS astrometry	
IMG_X		The x position of the object in the SKY image plane	
IMG_Y		The y position of the object in the SKY image plane	
NearestNeighbour	arcsec	The distance to the closest detection to this one, in this image.	
NearestOKNeighbour	arcsec	The distance to the closest Good or Reasonable detection to this one, in this image.	
DetFlag		The detection flag	
OpticalLoadingWarning	mag	Optical loading warning level	
StrayLightWarning		Whether this detection occurred within fitted stray light rings.	
NearBrightSourceWarning		Whether this detection occurred within the PSF wings of a fitted bright source	
MatchesKnownExtended		Whether the position of this source matches a known extended X-ray source.	
PileupFitted		Whether the accepted fit included pile up.	
SNR		The signal to noise ratio of the detection.	
CtsInPSFFit		Number of counts in the image region over which the final PSF fit was performed	
BGRateInPSFFit		Mean count rate in the background map in the region over which the final PSF fit was performed	
Cstat		$\mathcal{C}$ from the PSF fit	
Cstat_nosrc		$\mathcal{C}$ value if no source is fitted	
L_src		The likelihood value that this detection is not just a background fluctuation.	
Cstat_flat		$\mathcal{C}$ assuming a spatially uniform increase above the background	
Lflat		The likelihood value that this detection is PSF like, not flat	

Table 10 *continued*

Table 10 *(continued)*

Field	Units	Description	Has errors?
FracPix		The fraction of pixels within the PSF fit region which are exposed.	
Pileup_S		The best-fitting S parameter of the pile up model.	
Pileup_l		The best-fitting l parameter of the pile up model.	
Pileup_c		The best-fitting c parameter of the pile up model.	
Pileup_tau		The best-fitting tau parameter of the pile up model.	
Cstat_altPileup		$\mathcal{C}$ from the unusued fit. i.e. if the piled up model was used, this gives the Cstat from the non-piled-up fit, and vice-versa.	
PSF_Fit_Radius	pix	The radius of the circular region over which PSF fitting was carried out	
CellDetect_BoxWidth	pix	The full width of the cell-detect box in which this source was detected	
Rate	s^{-1}	The count rate of this detection	Yes
CtsInRate		The total number of counts in the region used to extract the count rate	
BGCtsInRate		The total number of counts in the region used to extract the count rate	
Rate_CF		The PSF correction factor for the count rate	
BGRateInRate	s^{-1}	The background rate in the region used for count rate calculation.	

Table 11. Contents of the 'Cross Correlations' catalog table, containing an entry for every match between a 2SXPS source and a source from another catalog

Field	Units	Description
2SXPS_ID		The 2SXPS sourceID
ExtCat_ID		The name of the source in the external catalog
Catalogue		The catalog containing the matched source
Distance	arcsec	The distance between the 1SXPS source and external catalog source
RA	degrees	The RA (J2000) of the source in the external catalog
Decl	degrees	The Declination (J2000) of the source in the external catalog
Err90	arcsec	The 90% confidence radial uncertainty in the external catalog position (inc systematics)