



**HAL**  
open science

## **STONKS: Quasi-real time XMM-Newton transient detection system**

E. Quintin, N. A. Webb, I. Georgantopoulos, M. Gupta, E. Kammoun, L. Michel, A. Schwope, H. Tranin, I. Traulsen

► **To cite this version:**

E. Quintin, N. A. Webb, I. Georgantopoulos, M. Gupta, E. Kammoun, et al.. STONKS: Quasi-real time XMM-Newton transient detection system. *Astronomy and Astrophysics - A&A*, 2024, 687, pp.A250. 10.1051/0004-6361/202348317 . hal-04660534

**HAL Id: hal-04660534**

**<https://hal.science/hal-04660534v1>**

Submitted on 24 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# STONKS: Quasi-real time *XMM-Newton* transient detection system<sup>★</sup>

E. Quintin<sup>1</sup>, N. A. Webb<sup>1</sup>, I. Georgantopoulos<sup>2</sup>, M. Gupta<sup>1</sup>, E. Kammoun<sup>1,3,4</sup>, L. Michel<sup>5</sup>,  
A. Schwobe<sup>6</sup>, H. Tranin<sup>7</sup>, and I. Traulsen<sup>6</sup>

<sup>1</sup> IRAP, Université de Toulouse, CNRS, UPS, CNES, 9 Avenue du Colonel Roche, BP 44346, 31028 Toulouse Cedex 4, France  
e-mail: [erwan.quintin@irap.omp.eu](mailto:erwan.quintin@irap.omp.eu)

<sup>2</sup> Institute for Astronomy Astrophysics Space Applications and Remote Sensing (IAASARS), National Observatory of Athens, I. Metaxa & V. Pavlou, Penteli 15236, Greece

<sup>3</sup> INAF – Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Firenze, Italy

<sup>4</sup> Dipartimento di Matematica e Fisica, Università Roma Tre, via della Vasca Navale 84, 00146 Rome, Italy

<sup>5</sup> Université de Strasbourg, CNRS, Observatoire Astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France

<sup>6</sup> Leibniz-Institut für Astrophysik, An der Sternwarte 16, 14482 Potsdam, Germany

<sup>7</sup> Institut de Ciències del Cosmos, Universitat de Barcelona, c. Martí i Franquès, 1, 08028, Barcelona, Spain

Received 19 October 2023 / Accepted 19 April 2024

## ABSTRACT

**Context.** Over recent decades, astronomy has entered the era of massive data and real-time surveys. This is improving the study of transient objects – although they still contain some of the most poorly understood phenomena in astrophysics, as it is inherently more difficult to obtain data to constrain the proposed models.

**Aims.** In order to help detect these objects in their brightest state and build synergies with multi-wavelength real-time surveys, we have built a quasi-real time automatic transient detection system for the *XMM-Newton* pipeline: the Search for Transient Objects in New detections using Known Sources (STONKS) pipeline.

**Methods.** STONKS detects long-term X-ray transient events by automatically comparing new *XMM-Newton* detections to any available archival X-ray data at this position, sending out an alert if the variability between observations (defined as the ratio between the maximum flux and the minimum flux or upper limit) is over 5. This required an initial careful cross-correlation and flux calibration of various X-ray catalogs from different observatories (*XMM-Newton*, *Chandra*, *Swift*, ROSAT, and eROSITA). A Bayesian framework was put into place to solve any ambiguous associations. We also systematically computed the *XMM-Newton* upper limits at the position of any X-ray source covered by the *XMM-Newton* observational footprint, even without any *XMM-Newton* counterpart. The behavior of STONKS was then tested on all 483 observations performed with imaging mode in 2021.

**Results.** Over the 2021 testing run, STONKS provided a daily alert rate of  $0.7^{+0.7}_{-0.5}$  alerts per day, about 80% of them corresponding to serendipitous sources. Among the detected variable serendipitous sources, there are: several highly variable active galactic nuclei (AGNs) and flaring stars, as well as new X-ray binary and ultra-luminous X-ray source candidates, some of which are present here. STONKS also detected targeted tidal disruption events, ensuring its ability to detect other serendipitous events. As a byproduct of our method, the archival multi-instrument catalog contains about one million X-ray sources, with 15% of them involving several catalogs and 60% of them having *XMM-Newton* (pointed or slew) upper limits.

**Conclusions.** STONKS demonstrates a great potential for revealing future serendipitous transient X-ray sources, providing the community with the ability to follow-up on these objects a few days after their detection with the goal of obtaining a better understanding of their nature. The underlying multi-instrument archival X-ray catalog will be made available to the community and kept up to date with future X-ray data releases.

**Key words.** methods: observational – methods: statistical – astronomical databases: miscellaneous – catalogs – X-rays: general

## 1. Introduction

The last few decades in the field of astronomy have witnessed a marked evolution in observational methods. More and more missions have turned toward time-domain astronomy, with large frameworks aimed at performing rapid follow-ups on transient events: among them, *Zwicky* Transient Facility (ZTF; [Bellm 2014](#)), SVOM mission ([Atteia et al. 2022](#)), *Vera C. Rubin* Observatory ([Ivezić et al. 2019](#)), and others. These missions often make

use of extremely large fields of view and high return rates aimed at achieving the greatest chance for detecting a transient event.

Because of the scarcity of X-ray photons and the need to be above the atmosphere to detect them, such all-sky monitorings have been significantly more difficult to implement in X-rays than in lower energies. Most of the current X-ray telescopes (with the exception of eROSITA and the upcoming *Einstein Probe*) instead perform observations of chosen targets, placed at the center of a relatively limited field of view of a few dozen square arcminutes, with typical exposure times ranging from a few to a few hundreds of kiloseconds. Within this field of view, a number of sources will be detected that are not the target and immediate subject of the observation; these detections are

<sup>★</sup> The multi-mission X-ray catalog is available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](https://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/687/A250>

referred to as “serendipitous” (typically  $\sim 75$  per observation for *XMM-Newton*, e.g., Webb et al. 2020). For most X-ray observatories, a significant effort has been put into detecting, filtering, and archiving these serendipitous sources, for which the various properties are generally summarized in the form of a catalog of detections (more details in Sect. 2.1.1).

The available X-ray catalogs contain hundreds of thousands of detections that cover many regions of interest over several decades. Systematically exploiting them is one of the current challenges of modern X-ray astronomy. One way to make use of these catalogs is to perform a classification of the sources, either by association with other catalogs (for instance Pineau et al. 2011) or by using more advanced probabilistic techniques (for instance Tranin et al. 2022). Once the sources are classified, it is possible to focus on a specific type of sources and thus provide an X-ray-selected population study of these objects (e.g., Vagnetti et al. 2011 for AGNs, Song et al. 2020 or Gúrpide et al. 2021 for ultraluminous X-ray sources, or Freund et al. 2022 for stars).

As it gives us access to more energetic events that are often intrinsically variable, the X-ray sky is even richer in transient events than the optical sky (e.g., Li et al. 2022). In the following paragraphs, we mention some instances of these sources and justify the interest of increasing their respective available samples.

Tidal disruption events (TDEs; e.g., Gezari 2021) correspond to the disruption of a star passing within the tidal radius of a black hole due to the strength of the tidal forces; this disruption can be either complete or only partial. The typical expected behavior is a sudden rise in the emission of the black hole, well described by a thermal continuum, followed by a slow decay over a few years, consistent more or less with a  $t^{-5/3}$  power-law decay (Rees 1988), or  $t^{-9/4}$  for partial TDEs (e.g., Coughlin & Nixon 2019). Surveys such as the ZTF (Bellm 2014) or the All Sky Automated Survey for SuperNovae (ASAS-SN; Kochanek et al. 2017) have allowed for the detection of dozens of optical TDEs (e.g., Hammerstein et al. 2022), while X-ray detected TDEs remain rare (e.g., Saxton et al. 2021). A comprehensive list of all TDE candidates can be found in the Open TDE catalog<sup>1</sup>. A large delay between the X-ray and optical counterpart of a TDE, as seen in ATLAS17jrp (Wang et al. 2022b), could explain the observational discrepancies (as any X-ray follow-up might be too early to catch the delayed X-ray counterpart to the initial optical event). Many questions remain unanswered about the precise emission mechanisms and the multi-wavelength counterparts of these events (Saxton et al. 2018). Two main points of interest about TDEs could justify the efforts of trying to find new candidates. The first advantage of TDEs is in the case of wandering IMBHs; outside of the massive flare due to the disruption of the star or a lucky lensing event, these black holes are practically undetectable. Observing TDEs in such environments is thus one of the preferred strategies for the detection of the still elusive IMBHs. The second point of interest in detecting TDEs is that the level of accretion reached during the flare goes well above the Eddington limit (Wu et al. 2018); the precise processes of super-Eddington accretion are still poorly understood, meaning that new samples of such processes could help us understand them.

A recently discovered phenomenon that seems to be linked to TDEs are quasi-periodic eruptions (QPEs), first discovered in 2019 (Miniutti et al. 2019) in a past X-ray TDE (GSN 069, e.g., Saxton et al. 2011; Shu et al. 2018). QPEs appear as large

$\sim 1$ h long outbursts of soft thermal X-rays, repeated every  $\sim 2$ – $10$  h, with peak luminosities of  $\approx 10^{42}$ – $10^{43}$  erg s<sup>-1</sup>. Only six QPE sources are known to this date: GSN 069, RX J1301.9+2747 (Sun et al. 2013; Giustini et al. 2020), eRO-QPE1 and eRO-QPE2 (Arcodia et al. 2021), along with two additional candidates, XMMSL1 J024916.6-041244 (Chakraborty et al. 2021), and Tormund (Quintin et al. 2023). Most sources have shown a pattern in their bursts, with large and small peaks alternating; eRO-QPE1 showed a transition from such a regular pattern to a chaotic profile with overlapping peaks in less than a week (Arcodia et al. 2022). The long-term evolution of GSN 069 is arguably the best constrained, with an overall decay of the emission over time, the bursts appearing only in a relatively low-flux state; a rebrightening was then observed, with the QPEs disappearing (Miniutti et al. 2023b). This was followed by a new decaying phase, and the QPEs appearing again, with a different alternating pattern than before (Miniutti et al. 2023a). Out of the six known QPE sources, three show a link with a past TDE (GSN 069, XMMSL1 J024916.6-041244, and Tormund). The precise emission mechanisms at play in QPEs are still unclear. Most models invoke either specific hydrodynamical instabilities (e.g., Sniegowska et al. 2020; Kaur et al. 2023; Pan et al. 2022; Śniegowska et al. 2023), repeated partial tidal disruption events (e.g., King 2020, 2022; Zhao et al. 2022; Wang et al. 2022a; Chen et al. 2022), or an initial partial TDE followed by repeated interactions between the remnant and its orbiting debris (e.g., Xian et al. 2021; Linial & Metzger 2023; Franchini et al. 2023). To discriminate between these models, more data are needed to both constrain the long-term evolution on the already-known QPE sources and to increase the sample of known QPE sources. This will allow us, for instance, to make statistically significant population studies (e.g., Wevers et al. 2022).

Another window on super-Eddington accretion is ultraluminous X-ray sources (ULXs; Kaaret et al. 2017). They correspond to extra-galactic, extra-nuclear sources reaching X-ray luminosities above  $3 \times 10^{39}$  erg s<sup>-1</sup>. This somewhat arbitrary threshold was chosen as it corresponds to the isotropic Eddington luminosity of a  $20 M_{\odot}$  black hole (Remillard & McClintock 2006). Going significantly above this value means that the source is either more massive than  $20 M_{\odot}$ , so that the Eddington limit can be respected. Otherwise, it violates this limit, which means that the accretion is following a super-Eddington regime. The discovery of accelerating coherent pulsations in a ULX in M82 (Bachetti et al. 2014) lead to the conclusion that at least some ULXs are host to a neutron star, and thus require highly super-Eddington accretion to reach the observed luminosities (up to  $500 L_{\text{Edd}}$  for the pulsating ULX in NGC 5907 reported in Israel et al. (2017) for instance). So far, only a handful of pulsating ULXs have been found. A key feature of these known PULXs is that they seem brighter and more variable than the overall ULX population, which could hint at a physically motivated sub-classification of ULXs, or be a selection bias due to the difficulty of finding pulsations in scarce X-ray signals. Nonetheless, outstanding variability has been used as a proxy to find good candidates for pulsations (Song et al. 2020) and could allow us to detect new candidates for further pulsation search.

While the previously mentioned variable sources are extra-galactic, our Galaxy is also rich in X-ray transient objects. For instance, some stars can be bright in X-rays (e.g., young stellar objects, Preibisch et al. 2005). Among these X-ray bright stars, some can show flaring episodes, which can be due to coronal activity for instance (e.g., Pallavicini et al. 1981), or to magnetic activity (e.g., Stelzer et al. 2013). These flares typically last for a few hours with peak luminosities in the  $10^{29}$ – $10^{32}$  erg s<sup>-1</sup> range

<sup>1</sup> <https://tde.space/>

**Table 1.** Properties of the catalogs after quality filtering.

Telescope	Catalog	Sky coverage (sq. degrees)	Limiting sensitivity ( $\text{erg s}^{-1} \text{cm}^{-2}$ )	Spatial resolution ( $FWHM$ , arcsec)	Sources	Detections	Dates	Reference
<i>XMM-Newton</i>	4XMMDR11	560	$\sim 10^{-15}$	5	470 000	700 000	2000–2020	Webb et al. (2020)
	4XMMDR11s	560	$\sim 10^{-15}$	5	34 000+	51 000+	2000–2020	Traulsen et al. (2019)
	XMMSL2	65 000	$\sim 10^{-12}$	10	22 000	27 000	2001–2014	Saxton et al. (2008)
<i>Swift</i>	2SXPS	3 790	$\sim 10^{-13}$	6	145 000	300 000	2005–2018	Evans et al. (2020b)
<i>Chandra</i>	CSC 2.0	550	$\sim 10^{-16}$	0.75–5	200 000	300 000	2000–2014	Evans et al. (2020a)
ROSAT	RASS	41 000	$\sim 10^{-12}$	20	60 000	60 000	1990–1991	Boller et al. (2016)
	WGACAT	7500	$\sim 10^{-13}$	20	70 000	80 000	1991–1994	White et al. (1994)
eROSITA	eFEDS	140	$\sim 10^{-14}$	5	20 000	20 000	Nov. 2019	Salvato et al. (2022)

**Notes.** The limiting sensitivities are typical flux values in the corresponding instrument’s energy band (see Fig. 3), but numerous instrumental effects (off-axis angle, background, exposure time) will impact this value. For *Chandra*, the two values for spatial resolution correspond to the on-axis and  $10'$  off-axis  $FWHM$ . For the *XMM-Newton* Stacked catalog, we only show the number of new sources and number of new detections (which might be associated with already known sources).

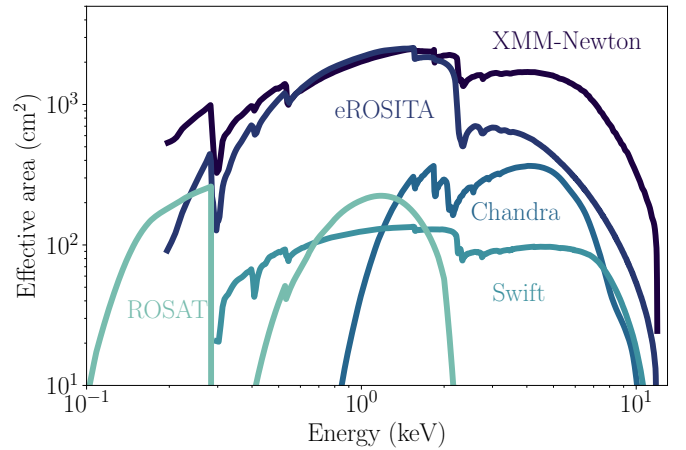
and are thus visible within observations of X-ray missions such as *XMM-Newton* (e.g., Pye et al. 2015, for a sample study).

On top of TDEs, QPEs, ULXs, and stellar flares, there is a host of other interesting X-ray variable sources: gamma ray bursts, novae (e.g., König et al. 2022b), cataclysmic variables (e.g., Webb et al. 2018), and X-ray binaries, supernovae, blazars, and changing-look active galactic nuclei (e.g., Graham et al. 2020). For all these events, an alert (and subsequent follow-up) even a week after the initial event can provide valuable information. Additionally, some newly studied variable sources are detected in other wavelengths and studying their possible X-ray counterparts might allow us to reveal or at least constrain their still unclear physical nature: fast blue optical transients (Margutti et al. 2019) and fast radio bursts (Petroff et al. 2019). Finally, there might even be new types of variable unknown X-ray objects lingering in the archives that are yet to be discovered.

All of these sources are rare and show some type of variability, either in flux or spectral shape. Finding and studying them would increase their numbers and help elucidate the underlying physical mechanism governing their nature. To improve our understanding of these sources, it thus seems profitable to find new candidates, based on X-ray variability. To be able to retrieve the most constraining data for these sources, both in X-rays and in other wavelengths, it is of paramount importance to detect them when they are in their brightest state.

In this paper, we describe a new quasi-real time transient detection system that could be deployed in the *XMM-Newton* pipeline, developed as part of the XMM2Athena project (Webb et al. 2023). Our approach is to compare new *XMM-Newton* EPIC detections to any available archival X-ray data, in order to assess the long-term variability of the underlying object. To do this in a computationally efficient manner that would not slow down the already-existing data stream, we performed a compilation of the archival X-ray sky (through both catalogs of detections and upper-limits). This catalog-oriented approach, on top of allowing for faster computations in the pipeline, also enables various data mining endeavours in the compiled X-ray archive, the results of which have been presented in earlier publications (e.g., Quintin et al. 2021, 2023).

We explain the underlying multi-instrument archival catalog and archival *XMM-Newton* upper limits (Sect. 2), then describe and test the proposed transient detection system itself (Sect. 3), and finally discuss the main limits, expected results and future updates of this system (Sect. 4).



**Fig. 1.** Comparison of effective areas of all the X-ray missions used in this work. For *XMM-Newton* we show the combined EPIC effective area. For *Chandra*, we show the ACIS-I effective area as of 2022. For *Swift* we show the XRT effective area. For eROSITA we show the effective area from the combined seven telescopes. For ROSAT we show the PSPC effective area.

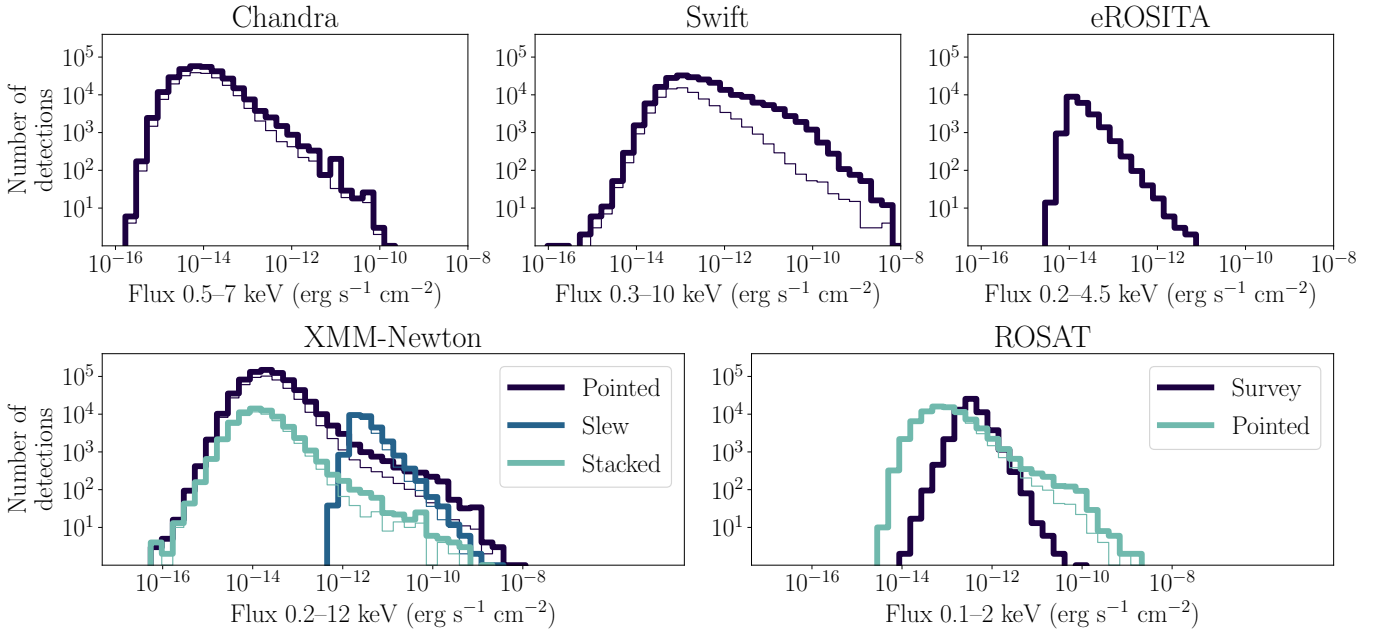
## 2. Collating the archival X-ray sky

### 2.1. X-ray multi-instrument matching method

#### 2.1.1. Data selection

Some studies have been performed to systematically look for variable objects in the archive of some X-ray observatories (e.g., the search for fast X-ray transients in the *Chandra* archive or the EXTraS project for *XMM-Newton*; Jonker et al. 2013; Luca et al. 2021). However, in order to improve our chances of finding long-term variability in serendipitous sources, a multi-instrument approach is preferable, as it provides an increased number of data points for a given source. For this reason, we used eight different X-ray catalogs, with complementary strengths and weaknesses. This method is similar for instance to the HILIGT web service (Saxton et al. 2022; König et al. 2022a). A summary of the catalogs’ respective properties can be found in Table 1 and their effective areas are shown in Fig. 1.

The first three catalogs we chose are 4XMM DR11 (Webb et al. 2020), 2SXPS (Evans et al. 2020b), and 2CXO



**Fig. 2.** Flux distributions of each X-ray observatory used in this study, in their native energy band, with the different catalogs shown in different colors. For each catalog, we show the flux distribution of all detections (thick line), as well as the flux distribution averaged for each source (thin line). The difference between the detection-wise and source-wise flux distributions depends on the observational strategy of each X-ray instrument.

(Evans et al. 2020a), which are the source catalog respectively for *XMM-Newton*, *Swift/XRT*, and *Chandra*. Their respective sensitivity, angular resolution and sky coverage (see Table 1) differ significantly because of the different technical setups of their instrumentation, driven by different scientific goals.

We also took into account two additional catalogs obtained from *XMM-Newton*: the slew catalog XMMSL2 (Saxton et al. 2008) and the stacked catalog 4XMM DR11 Stacked (Traulsen et al. 2019). The first one corresponds to detections obtained during the slewing of the instrument, between two consecutive pointings. It provides us with a large sky coverage, at low exposure times and thus low sensitivity. The second catalog is obtained from the stacking of overlapping observations, which provides improved sensitivity and more reliable source parameters compared to single observations, as well as possibly new detections in some observations. For the stacked catalog, we only kept detections that were not in the initial pointed catalog (corresponding either to sources that are in the initial catalog but for which some observations did not lead to clean detections, and also for entirely new sources absent from the initial catalog).

We added two ROSAT catalogs, 2RXS (Boller et al. 2016) and WGACAT (White et al. 1994), corresponding respectively to the sky survey and to subsequent pointed observations. Despite their relatively low sensitivity and angular resolution, these catalogs are very useful for their wide sky coverage, as well as for the fact that they provide us with a longer temporal baseline to study variability.

Finally, the study of long-term variability of X-ray sources will be immensely improved by the data from eROSITA (Predehl et al. 2021), which will provide multiple all-sky X-ray surveys with sensitivity levels comparable to that of *XMM-Newton*. In order to make a proof of concept of the interest of using future eROSITA data within our framework, we have used the available early data from the eROSITA Final Equatorial Depth Survey catalog (eFEDS; Salvato et al. 2022), which covers a small patch of the sky of about 140 square degrees, with non-contemporaneous *XMM-Newton* and *Chandra*

observations. Boller et al. (2022) have already performed a study of the variable sources in eFEDS, although our method should reveal additional long-term variability.

Once selected, these catalogs have been cleaned using different selection criteria with the aim of keeping only point-like sources, avoiding spurious detections and improving the overall quality of the final catalog. The cleaning procedures were performed on detections; the remaining sources are those that have at least one remaining clean detection. The various catalog-specific selection criteria are summarized in Appendix A.

The resulting flux distributions of each catalog are shown in Fig. 2. In particular, this figure shows the flux distribution of all detections, as well as the flux distribution averaged for each source. The shape of these distributions and the differences between them will depend on the overall observing strategy – for instance, the *Swift* flux distribution loses a significant fraction of its high-flux component when averaging over each source, because *Swift* is often used as a monitoring telescope for bright objects.

Once these quality checks have been applied, we have a total of about 1 million X-ray catalog sources and 1.5 million detections. For each detection, we have a rate in the corresponding total energy band of the instrument, as well as in different subbands that will be used to access spectral information. We now need to associate those sources together. This will be done by matching the catalogs two by two at first, in order to take into account their respective astrometric differences and avoid the combinatorial difficulties of a single multi-catalog match; then, these two-by-two matches will be combined into multi-catalog sources, using a conservative fusion approach.

### 2.1.2. Two-by-two catalog matches

The core of our method is based on the two-by-two correlations between catalogs. These were performed using STILTS (Taylor 2006), based on the positions and  $3\sigma$  circular position errors for each source in the two considered catalogs. Among

all combinations of catalogs, we did not compute the *XMM-Newton* pointed to *XMM-Newton* stacked cross-correlation, as this work was already performed and manually screened in the elaboration of the *XMM-Newton* stacked catalog (Traulsen et al. 2019). Two issues arose from this naive cross-matching method.

The first issue we encountered was for very bright X-ray sources ( $F \sim 10^{-10}$  erg s $^{-1}$ ). For these sources, the large number of photons allowed for a very precise fit of the PSF; so precise in fact that the  $3\sigma$  positional errors can be smaller than the astrometric error between catalogs, thus preventing the matches for bright sources. To prevent this, we have computed an estimation of the astrometric error for each catalog combination, by producing a naive correlation and taking the closest match for each source using a very large position cutoff (1 arcmin). Assuming that the coordinate differences follow the same normal distribution, the angular distance distribution of this naive match should yield a Rayleigh distribution at close distance, with an excess at large distance due to spurious associations (this method was used for instance in Boller et al. 2016). Taking the maximum of this Rayleigh distribution allows us to retrieve its  $\sigma$  value, which roughly corresponds to the standard deviation of the coordinate errors. For the ulterior matches between those two given catalogs, the matching distance was taken as the maximum between the  $3\sigma$  position error and the estimated astrometric error.

The second issue arises for ambiguous correlations. Indeed, taking the  $3\sigma$  positional error and the astrometric error into account can lead to a reasonably large maximum matching distance, that can then lead to a number of possible counterparts. In this case, the STILTS command will return a group of ambiguous associations, with all allowed combinations of source associations. Identifying the correct counterpart for each source is essential, as spurious associations may lead to large, erroneous variability. For this purpose, we have developed a Bayesian approach to quantify the quality of an association, which will allow us to compare between candidates and decide whether the match is decisive or unclear. The precise method is similar to the one implemented in NWAY (Salvato et al. 2018), which was inspired from Budavári & Szalay (2008). We denote  $H_i$  as the hypothesis that the  $i$ th possible match between two catalog sources is real, and  $\bar{H}_i$  as the opposite hypothesis; the data, namely, the position and position error of each source, are noted as  $D_i$ . The Bayesian probability for the  $i$ th match is thus:

$$P(H_i|D_i) = P(D_i|H_i) \times \frac{P(H_i)}{P(D_i)}. \quad (1)$$

The end goal will be to compute the ratio of this value between different counterparts,  $i$ . With a flat prior on the data and  $P(H_i)$  only depending on the overlap between two catalogs and thus independent of  $i$ , for a given catalog combination the only value of interest is  $P(D_i|H_i)$ . With the same assumptions as the Appendix B from Budavári & Szalay (2008) (i.e., a spherical normal error on position, with error bars and distances small compared to the size of the sky), this value is given by:

$$P(D_i|H_i) = \frac{2}{\sigma_1^2 + \sigma_2^2} \exp\left(-\frac{\psi^2}{2(\sigma_1^2 + \sigma_2^2)}\right), \quad (2)$$

with  $\sigma_1$  and  $\sigma_2$  the error bars of the two associated sources and  $\psi$  the angular distance between their positions; at this stage, the astrometric error is not taken into account. We compute this ‘‘association score’’ for all associations, and use it as a way to compare between ambiguous ones. After manual screening, we take a ratio of 3 between two scores as a very good indication

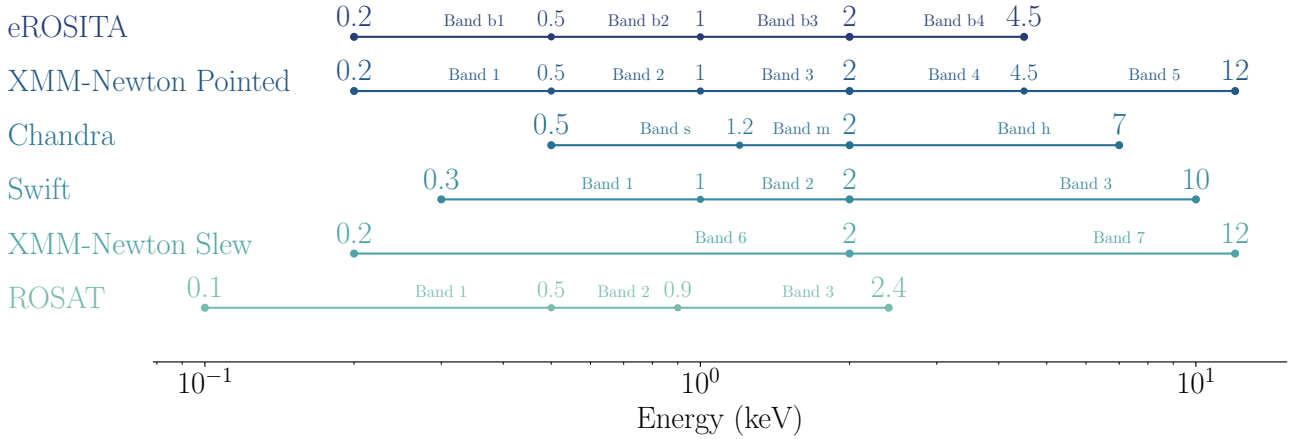
that one association is favored over the other; a ratio below that generally corresponds to different spatial resolutions resulting in two sources for an instrument being seen as a single source for another instrument (*Chandra* vs. *Swift* typically).

The precise workflow for each two-by-two catalog correlation is thus as follows: we first estimate the astrometry error between two catalogs by performing a crude correlation, and taking its typical angular distance; we perform the precise correlation using  $3\sigma$  positional errors and astrometric error; the association score for all associations is computed following Eq. (2). Then, for each group of ambiguous associations, we sort by order of association score. We compare the score of the most probable association of the group to the score of the second most probable association involving any of the two concerned sources (this is the major difference with NWAY, in which only the possible matches for one source of the pair are considered). If the ratio is higher than 3, we validate the first association and ignore all the other ones; else, we ignore all the associations for these two sources, as it is impossible to safely conclude on the association. Finally, we proceed until all combinations have been either accepted or ignored.

Deviating from Budavári & Szalay (2008), we do not include photometric information in our Bayesian approach, because a photometry-based match relies on constant flux assumption, while we search for transients. One issue that may arise from this choice is to favor a close spatial match between a bright and a faint source from two catalogs, where one of them has poorer spatial localisation (e.g., ROSAT or *XMM-Newton* slew), while the correct bright (non-variable) match is not favored spatially. This can be avoided by using the ambiguous match solver, which will be able to flag such situations. This can also be manually treated at the quality check step (see Sect. 3).

### 2.1.3. Combined catalog matches

Once all two-by-two correlations of catalogs are performed, we need to merge these into multi-catalog associations. This requires dealing with associations that are inconsistent between catalogs. We chose a conservative approach, in which chain-like correlations are refused (i.e., with three sources from catalogs A, B, and C, source B is associated with both A and C, but A and C are only associated with B and not with each other). To do this, we first classify the catalogs in an arbitrary order of interest, with the idea that such chains will be dealt with in order of priority (i.e., sources A and B first in the previous example). In a pair of catalogs, the first is hereafter called primary, the other secondary. We compute all two-by-two correlations for the primary catalog with any secondary catalog, including solving ambiguous correlations using the association score, as presented in the previous section. For each source from the primary catalog, we validate its associations with all its corresponding secondary sources into the final multi-instrument catalog. At this stage, we should have recovered any counterpart to each source of the primary catalog. We then reiterate this procedure by promoting the secondary catalog to primary. However, an additional condition to accept an association now is that neither the (new) primary, nor the secondary sources, have already been encountered at a previous stage in this procedure. If they had already been encountered, this means that they are either already part of a validated association, or part of a chain-like association, which is prohibited. We proceed with this, until all two-by-two catalog correlations are merged into a single multi-catalogs catalog, where associations are performed both conservatively and quantitatively, through the use of the Bayesian association score.



**Fig. 3.** Energy bands of the various catalogs and instruments used in this work. We also show the catalog-specific internal energy bands, with their catalog name indicated above their respective energy regime.

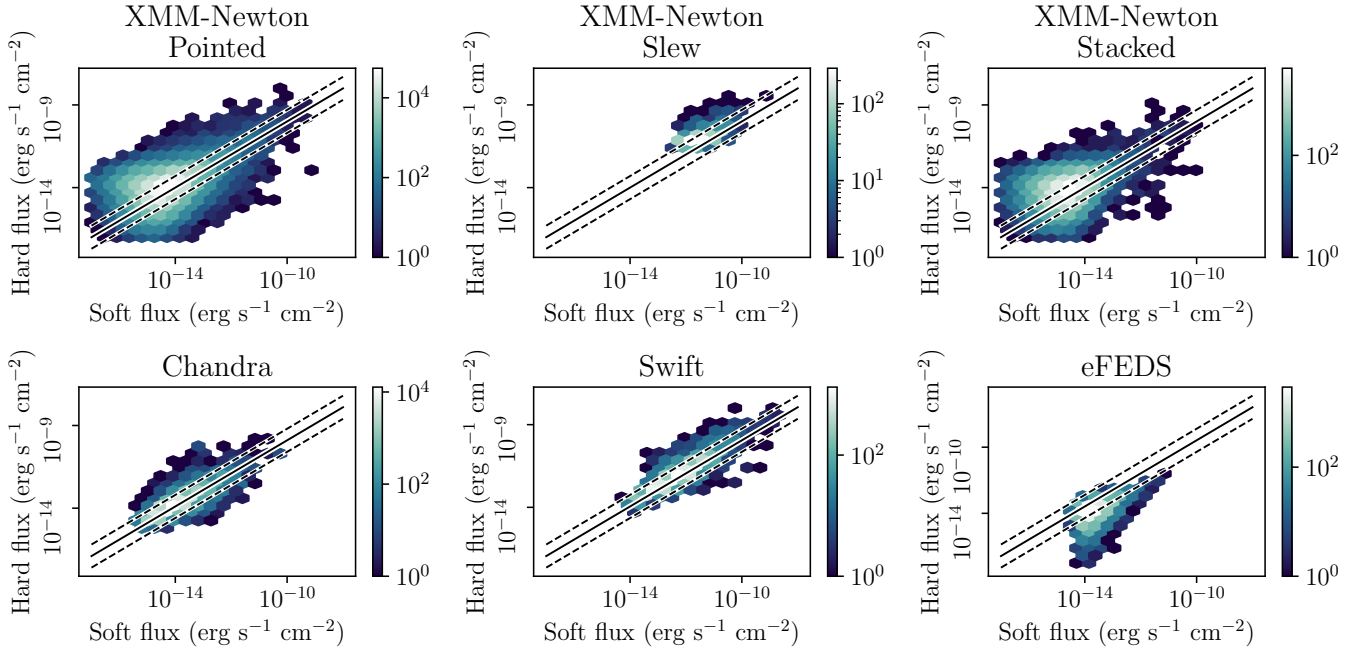
## 2.2. Cross calibration

Once sources are associated in the multi-instrument catalog, we need to compare the various fluxes of each catalog source. However, reliable cross-calibration of the various instruments is a major challenge for any multi-catalog flux comparison. Each instrument has a different response (see Fig. 1). While most of those instrumental effects are taken into account by the processing pipelines through ancillary and response files, some biases remain (of the order  $\sim 8\%$  between the EPIC instruments for instance, [Smith 2022](#)), and about 5–15% between different missions when working in the same energy band (e.g., [Madsen et al. 2017](#)). However, the energy bands differ between the missions. Figure 3 shows the respective total energy bands of each specific catalog, as well as the catalog-dependent internal energy bands. A useful feature one can see in this figure is that, for all catalogs, the value of 2 keV is a limit to some internal bands.

To compare the fluxes obtained by different instruments and assess the source’s variability, we first need to convert each detection to a single, common energy band; we cannot directly compare for instance the *XMM-Newton* flux of a source in the 0.2–12 keV band, to that of *Chandra*, which is optimised in the 0.5–7 keV band. The common band we chose to compute fluxes is the 0.1–12 keV band, as it allows us to constrain the energy bands of every one of the missions we used (*XMM-Newton* going to the highest energies and ROSAT to the lowest). Then, to extrapolate the instrument detections to this common band, we need to assume a specific spectral shape. We chose an absorbed power-law, of parameters  $\Gamma = 1.7$  and  $N_H = 3 \times 10^{20} \text{ cm}^{-2}$ . The reason this was chosen is that these parameters correspond to a typical X-ray source (e.g., [Watson et al. 2009](#)), and the resulting spectrum is thus rarely far from the actual spectrum of the source – for this reason, it was used to compute fluxes for instance in the *XMM-Newton* and *Swift* catalogs. Any other spectral model would not be self-consistent with the direct use of the catalog fluxes (which use this assumption), and would thus require further calibration. Assuming this fixed spectral shape, the contributions to the total flux of each band as well as the fraction of the flux missed by each instrument is shown in Table B.1.

This spectral shape assumption has its limits. It fits relatively well to the majority of sources, however, for the softest or hardest sources there can be some discrepancy. Figure 4 gives the distribution of the soft vs. hard fluxes ( $<2\text{keV}$  vs.  $>2\text{keV}$ ) for

each detection in the instruments with a hard energy band (i.e., not ROSAT). Any departure from the black line means a departure from the assumed spectral model. To validate the use of this spectral assumption in order to assess variability between detections of different instruments, it is necessary to estimate the spurious variability that would appear from wrongfully extrapolating the source’s flux beyond the specific instrumental bands. For this purpose, we implement two tests. The first test of validity of our spectral assumption simply consists in computing the error in flux estimation arising from this assumption, depending on the source’s true spectral shape. In practice, we compute the evolution of the extrapolated flux from each mission’s band to the total band assuming a fixed  $\Gamma = 1.7$  and  $n_H = 3 \times 10^{20} \text{ cm}^{-2}$ , depending on the actual photon index of the source (in a 0.5–4 range). A photon index of  $\sim 4$  is reasonably close to a soft thermal emission, at least from a catalog point of view. The various fluxes were computed using JAXspec ([Barret & Dupourqué 2024](#), Dupourqué et al., in prep.). The results can be seen in Fig. 5. In this figure, one can see that in this range of photon indices, while the spectral assumption indeed leads to a bias on the estimated flux, this bias stays overall below a factor of five. More importantly, the respective biases of different missions stay closer than a factor of five from each other, which means that at a given value of  $\Gamma$ , the calibration method should lead to a minimal number of spurious alerts. To assess the effect of such extrapolation on data rather than theoretical spectra, we test it on the *XMM-Newton* data, and analyse the variability that is created solely from this method. We started by truncating the energy bands of *XMM-Newton* to fit those of *Chandra*, which is the second most delicate extrapolation after ROSAT. For each *XMM-Newton* detection, we removed the first and last bands, to retrieve *XMM-Newton* fluxes in the 0.5–4.5 keV. To get the same higher energy limit, namely, 7 keV for *Chandra*, we had to extrapolate the flux of the *XMM-Newton* band 4 from 2–4.5 keV to 2–7 keV. This extrapolation is done using the spectral assumption of an absorbed powerlaw, with the aforementioned parameters. The effect of this assumption on a single band is much smaller than on the entire *XMM-Newton* energy bandwidth, and is thus neglected. After this, we extrapolate the simulated 0.5–7 keV flux to the 0.1–12 keV band using the same conversion factor we would have used for *Chandra* data. Comparing the resulting flux to the actual 0.2–12 keV *XMM-Newton* detection allows us to assess the spurious variability caused by this spectral approximation (the 0.1–0.2 keV contribution is negligible in this spectral



**Fig. 4.** Comparison between the hard and soft fluxes for each mission with hard detections (i.e.,  $>2$  keV). The black lines show the expected behavior of the spectral assumption (absorbed power law of  $N_{\text{H}} = 3 \times 10^{20} \text{ cm}^{-2}$  and  $\Gamma = 1.7$ ), and the black dotted lines show a departure by a factor of 5 from this behavior. While the spread around the assumed shape can appear significant, it is important to remember that the error bars on these hard and soft fluxes are significant as well (typically signal to noise ratio of about 3 or less), so the statistical significance of the spread is reduced

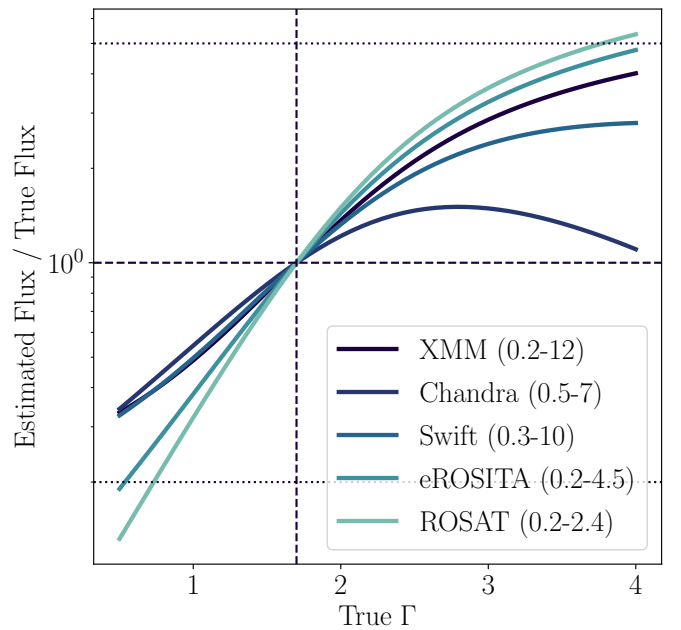
assumption). We use a conservative estimate of the variability between the two flux computations. We compute the ratio of the higher of them minus its error over the lower flux plus its error:

$$V_{\text{Conservative}} = \begin{cases} \max\left(\frac{F_{\text{Band 8}} - \sigma_{\text{Band 8}}}{F_{\text{Extrap.}} + \sigma_{\text{Extrap.}}}, 1\right) & \text{if } F_{\text{Band 8}} > F_{\text{Extrap.}} \\ \min\left(\frac{F_{\text{Band 8}} + \sigma_{\text{Band 8}}}{F_{\text{Extrap.}} - \sigma_{\text{Extrap.}}}, 1\right) & \text{if } F_{\text{Band 8}} < F_{\text{Extrap.}} \end{cases} \quad (3)$$

with  $F$  and  $\sigma$  the respective flux and  $1\sigma$  flux errors for both methods. This estimate takes the value of 1 in the case both methods are consistent at a  $1\sigma$  level, and otherwise takes the most pessimistic assumption for variability. This metric was used because it is similar to the one used later on for variability alerts (see Eq. (4)), a source being labeled as variable if this metric is above 5 (or here below 0.2 as well).

The resulting spurious variabilities can be seen in Fig. 6. We retrieved about 4000 spurious alerts out of the 700 000 detections, amounting to about 0.6% false alert rate. These alerts are indeed caused by the softest and hardest sources of the catalog, for which the assumption does not hold well – this can be verified in the right panel of Fig. 6, showing the difference in density distribution of hardness ratios of the false alert detections.

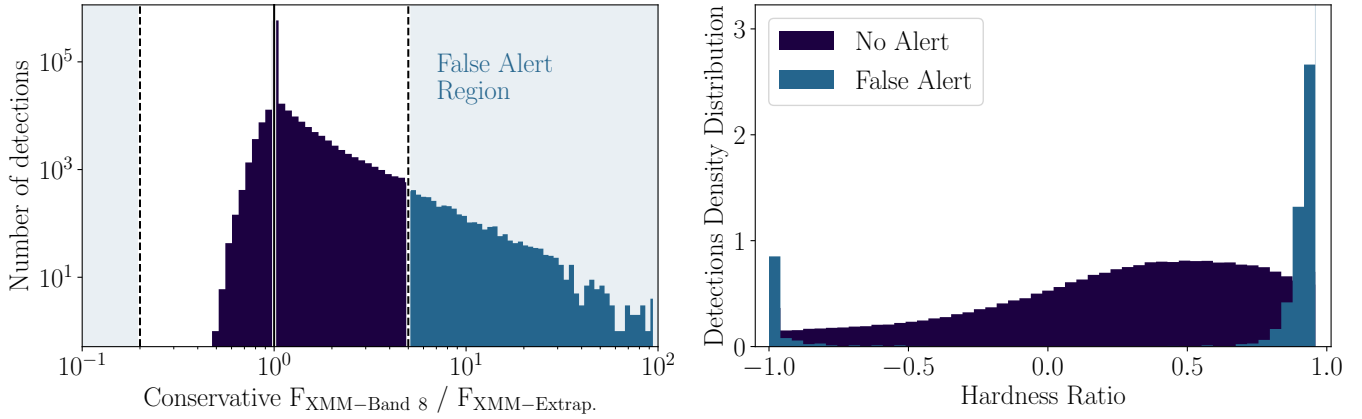
This spurious alert rate is reasonably small, however the total alert rate being about  $\sim 2.5\%$  of detections (see Sect. 3.2), this leads to a contamination of the alerts by at most  $\sim 20\%$  and could warrant further attention. While a more adaptive spectral approximation would be possible (e.g., based on the measured hardness ratio), this solution would be very biased for low signal-to-noise detections, that tend to be significantly harder or softer than bright detections purely because of statistical effects. This would in turn dramatically increase the false alarm rate for faint detections, which is not desirable. Additionally, a minority of detections from the multi-instrument archives have available



**Fig. 5.** Evolution of the ratio between the flux extrapolated from each mission band assuming  $\Gamma = 1.7$  and  $n_{\text{H}} = 3 \times 10^{20} \text{ cm}^{-2}$ , and the true flux of a source, depending on the value of its photon index  $\Gamma$ . The dashed lines correspond to the reference ( $\Gamma = 1.7$  and ratio of 1), and the dotted lines correspond to a factor of 5. While ROSAT goes over the threshold of 5 for the softest sources, what matters most to our study is that at a given  $\Gamma$  the ratio between different missions is below five (to avoid spurious alerts).

hardness information (e.g., only  $\sim 20\%$  of both the *Chandra* and *Swift* archives). Overall, proper spectral data is simply not widely available in a purely catalogue-oriented approach, and a common spectral assumption is justified (which is why this solution





**Fig. 6.** Assessment of the effect of the spectral assumption on variability estimates. *Left panel:* distribution of the conservative estimate of the variability between the true flux, and the one obtained after cropping to the *Chandra* bandwidth and extrapolation to the 0.1–12 keV band. All detections with a variability larger than a factor of 5 between both methods would lead to spurious transient alerts. *Right panel:* comparison between the hardness ratio density distributions of the detections that lead to spurious alerts (light blue) and the ones without alerts (dark blue). This confirms that spurious alerts can happen in the case where the spectral assumption does not fit the data well, that is, for extreme hardness ratios.

is already implemented for each respective catalog). Alternative methods for flux extrapolations, using additional data not present in the catalogs, will be explored in the future (e.g., using the archive-wide spectral fitting computed for *XMM-Newton* as part of XMM2Athena, Webb et al. 2023). For now, we put in place different safeguards to warn and help the user in the case of a possible failure of this assumption, presented in Sect. 3.

### 2.3. Upper limits

Correlating the sources from several catalogs allows us to retrieve the flux evolution of a given physical source between several epochs. The main use case of this method is when the source was detected in the different catalogs individually. However, this method also allows us to uncover valuable information in the case where it was observed but not detected by one of the instruments. Indeed, the fact that a source was within the field of view of a given observation but not detected means that it was, at the moment of the observation, below the sensitivity of the used detection method at this point of the instrument. By computing the said sensitivity, we can retrieve an upper limit on the source’s flux. This phenomenon takes place in two instances: either its intrinsic flux is constant and the observation in which it was detected previously has a better sensitivity than the one that missed it; or, the source is transient.

We put this idea into practice for the *XMM-Newton* upper limits. We selected two types of sources for the upper limits computation: the first type of sources are known, detected-at-least-once *XMM-Newton* sources. This allows us to check whether these known *XMM-Newton* sources were detected every time they were observed, which is a piece of information absent from the *XMM-Newton* base catalog, but present in the *XMM-Newton* stacked catalog. The second type of source for which the *XMM-Newton* upper limits are relevant are for the sources only present in other catalogs, but that have been observed by *XMM-Newton*. Using the 4XMM DR11 Multi-Order-Coverage map (MOC; Fernique et al. 2014) which provides us with the spatial footprint of the observations, we selected all multi-catalog sources that lie within this MOC but are far away ( $>10''$ ) from any *XMM-Newton* source. This was done using the MOCpy package (Boch 2019). For all those sources, the upper limits were computed using RapidXMM (Ruiz et al. 2022). We only kept

the upper limits with a 0.2–12 keV quality flag of 0, and that were not simultaneous with a *XMM-Newton* stacked detection. We then converted the obtained  $1\sigma$  count-rates upper limits to 0.2–12 keV flux upper limits, using the same spectral assumption of a power-law of photo-index  $\Gamma = 1.7$  and  $N_H = 3 \times 10^{20} \text{ cm}^{-2}$ . While the RapidXMM framework provides pre-computed upper limits for all three EPIC instruments individually, we used the mathematical framework presented in Ruiz et al. (2022) to compute the EPIC combined  $3\sigma$  flux upper limits, in order to obtain more constraining upper limits.

Additionally, we used upper limit information from both *Chandra* and *Swift*, but only for their respective sources. For *Chandra*, the non-detections of *Chandra* sources are directly available in the catalog. For *Swift*, the upper limits are not readily available in a catalog-based approach, but we have access to the stacked detections. They correspond to the average flux for a source over all its *Swift* exposures, and also provide us with the dates for the first and last observations. Thus, any *Swift* detection that is significantly above a stacked *Swift* flux hints at variability (for an example, see Figs. C.6 or C.7).

### 2.4. X-ray multi-instrument catalog properties

#### 2.4.1. Matching statistics

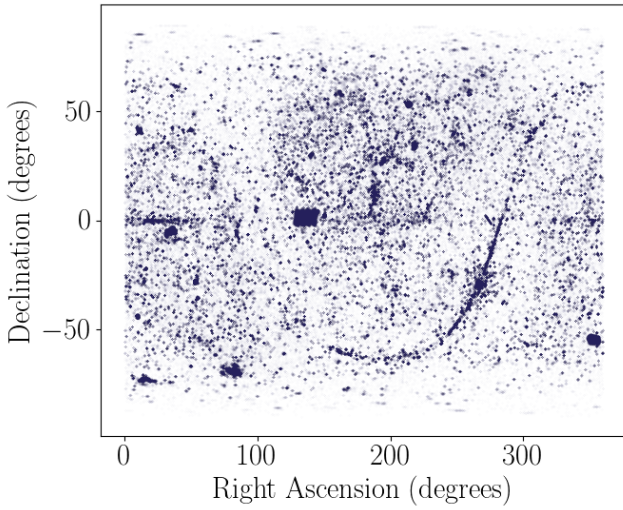
The cross-matched catalog consists of 926 753 multi-catalog sources, to be compared with the initial 1 258 420 single-catalog sources before the cross-match. Because of the sparse X-ray coverage of the sky (see Fig. 7 for a sky map of the final catalog), most of the final sources only contain data from one catalog, but the remaining 15% of the final sources (99 208) show multi-catalog data (see top panel in Fig. 8). The catalog-wise matching properties in terms of number and typical offsets are summarized in Table 2, and the distribution of number of catalogs per cross-matched source is shown in Fig. 8.

The underlying goal of this multi-catalog method was to increase the number of data points available per source, in order to be able to better estimate the underlying object’s variability. The catalog cross-matching allowed us to increase the average number of detections per source from 1.55 to 1.75. The use of upper limits allowed us to further increase the average number of data points (detections and upper limits combined) from 1.75

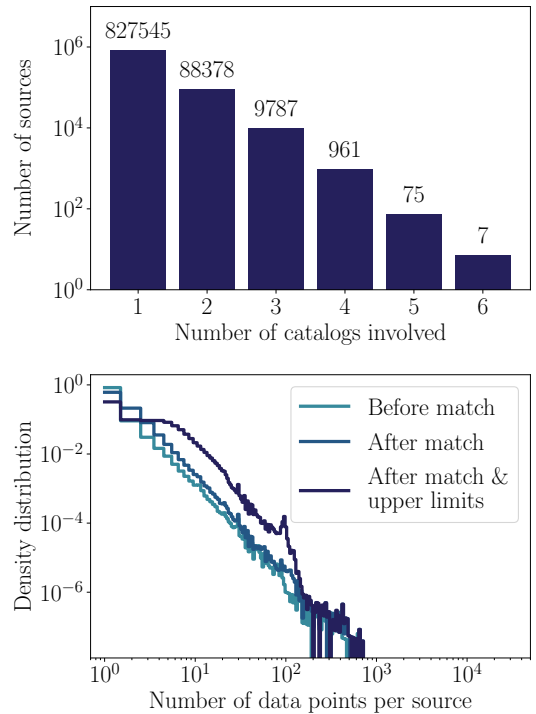
**Table 2.** Final two-by-two cross match statistics of our multi-instrument catalog.

Cross-match	<i>Chandra</i>	<i>Swift</i>	eFEDS	XMM slew	ROSAT survey	ROSAT pointed	XMM stacked (without pointed)
XMM pointed	48 106 1.4''	27 710 2.6''	1364 3.6''	1368 5.4''	1408 13.8''	6294 9.9''	N/A
<i>Chandra</i>		10 055 2.3''	177 3.2''	558 5.8''	619 13.3''	2472 11.6''	1537 1.1''
<i>Swift</i>			281 3.8''	3345 5.6''	4114 12.8''	3992 11.5''	343 2.3''
eFEDS				52 5.9''	148 14.3''	1 20.7''	34 3.2''
XMM slew					4690 12.9''	1721 17.8''	15 5.2''
ROSAT survey						3865 31.5''	14 14.3''
ROSAT pointed							77 10.2''

**Notes.** For each combination of catalogs, we show the number of final multi-instrument sources involving both the catalogs, as well as the median angular distance between these sources. As a reminder, we did not compute the *XMM-Newton* pointed to *XMM-Newton* stacked cross-correlation, as this work was already performed and manually screened in the elaboration of the *XMM-Newton* stacked catalog.

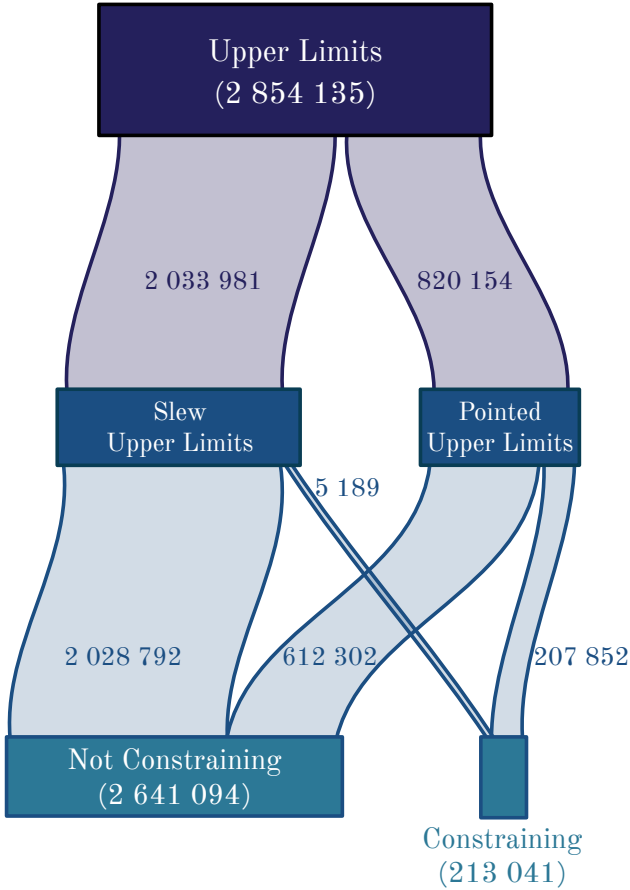


**Fig. 7.** Sky map of the multi-instrument catalog. The galactic plane is visible, as well as the eFEDS field of view around RA  $\sim 130^\circ$  and Dec  $\sim 0^\circ$ . This shows the inhomogeneity of the archival X-ray sky coverage.



**Fig. 8.** Illustration of the gain in information on the long-term evolution of X-ray sources, obtained thanks to the cross-matching & upper-limits. *Top panel:* distribution of the number of catalogs involved in each multi-catalog source. The majority of the sources only have data for one catalog, but for the remaining 15% at least two catalogs are involved. Despite using 7 catalogs, no source was detected in all of them (mostly due to the very constraining sky coverage of the eFEDS catalogs). *Bottom panel:* density distribution of the number of data points per source, before the cross-match in light blue, after the match in blue, and after taking into account upper limits in dark blue. Both the cross-match and the use of upper limits allows us to increase the number of data points per source, namely, skew this density distribution to the right.

to 5.0 (see precise statistics in the next section). The precise density distributions of the final number of data points per source is available in Fig. 8. In particular, the number of sources with only one detection (i.e., for which estimating the variability is impossible) went down from 839 361 to 675 829 thanks to the instrument cross-matching, and is further reduced to 302 252 once upper limits are taken into account. For sources which already had several available data points, the cross-matching allows us to improve the temporal coverage of the source, either by diminishing the average time between two consecutive data points, or by increasing the total coverage (i.e., time between the first and last available data points).

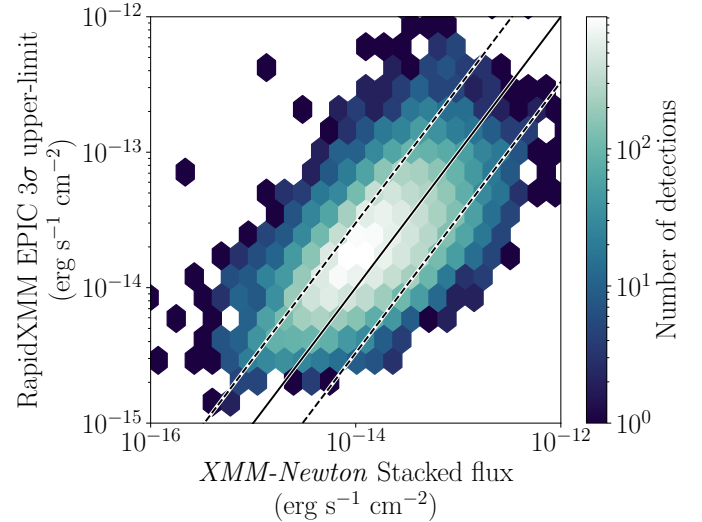


**Fig. 9.** Statistics for the 2 854 135 RapidXMM upper limits on multi-instrument sources in the 4XMM DR11 MOC. These combine the three EPIC instruments, and are 0.2–12 keV flux  $3\sigma$  upper limits. An upper limit is considered constraining if it is lower than the lowest flux value of the corresponding multi-instrument source. Most upper limits are from the slews of the catalog, although these are seldom constraining.

#### 2.4.2. Upper limits statistics

We called RapidXMM on the 586 483 multi-instrument sources that lie in the *XMM-Newton* MOC – out of those, 116 926 are not 4XMM DR11 sources. Half of these (65 939) are faint *Chandra* sources, and the rest are either *XMM-Newton* Stacked detections with no clean association in the normal catalog (31 628), *Swift* stacked detections, or some *XMM-Newton* slew transients or unflagged spurious detection (mostly in extended sources for which the *XMM-Newton* slew extent is falsely zero due to low counts).

The statistics of the resulting upper limits are shown in detail in Fig. 9. We retrieved 2 854 135 upper limits, 70% being *XMM-Newton* slew upper limits and 30% being for pointed observations. The overwhelming majority (92%) of these upper limits are not constraining, in the sense that they are higher than the lowest recorded flux of the corresponding multi-instrument source. However, for 213 041 upper limits (corresponding to 63 795 individual multi-instrument sources), they are indeed constraining, thus allowing us to improve our constraint on the variability of the underlying objects. Among these sources, 13 497 do not correspond to either a *XMM-Newton* pointed or stacked source, meaning that a multi-instrument approach was necessary in constraining the variability of the underlying objects.



**Fig. 10.** Comparison between the RapidXMM  $3\sigma$  0.2–12 keV flux upper limits, and the corresponding *XMM-Newton* stacked 0.2–12 keV flux detections. The black line shows a one-to-one behavior, and the dashed black lines show a departure by a factor of three from this behavior.

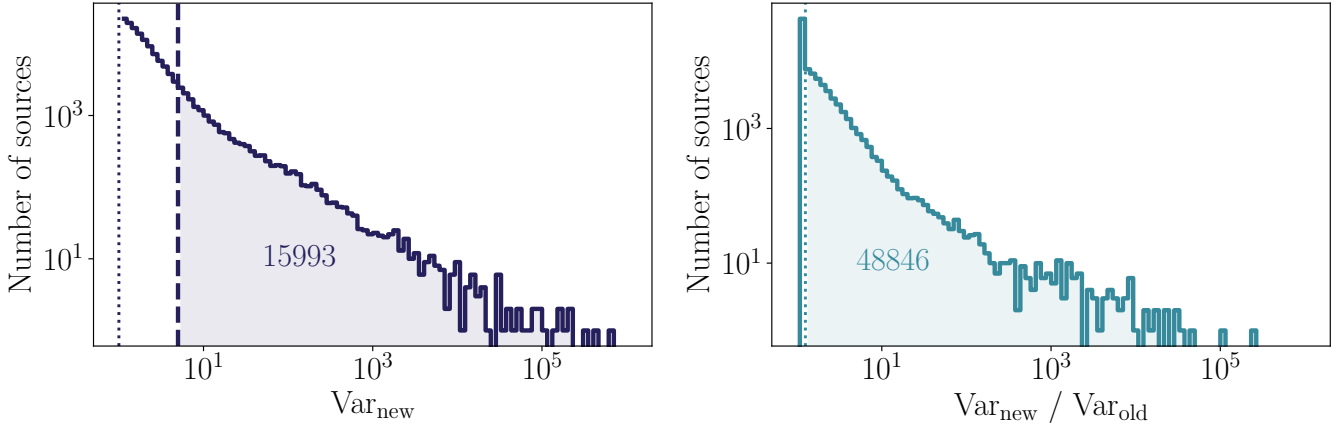
We chose not to use RapidXMM upper limits in the case where a flux value is available from the *XMM-Newton* stacked catalog, which provides measurements in all covering *XMM-Newton* observations. This was justified by the additional manual screening that the *XMM-Newton* stacked catalog went through. However, as a side result, we were able to assess the quality of the RapidXMM upper limits by comparing them to the simultaneous *XMM-Newton* stacked detections, which underwent several additional steps of screening. The resulting comparison between the 22 161 relevant detections is shown in Fig. 10. Overall, the majority (82%) of the RapidXMM  $3\sigma$  upper limits are within a factor of three of the corresponding *XMM-Newton* stacked detection. Once the *XMM-Newton* stacked flux error bars are taken into account, this fraction goes up to 99%, demonstrating coherence between the two methods. In particular, this confirms the quality of the RapidXMM flux constraints in the case where no *XMM-Newton* stacked source is present, that is, transients that were bright in another catalog.

#### 2.4.3. Variability statistics

After performing both the catalog cross-correlation and *XMM-Newton* upper limits computation, we obtain a large multi-instrument X-ray archival catalog. While such a tool can have various applications for data mining endeavours, systematically exploiting this catalog is beyond the scope of this work. However, we are particularly interested in one information, the long-term variability of sources. Among the various ways to define the variability of an object, we chose to use the pessimistic flux variability amplitude:

$$V = \frac{\max(F_{\text{low}})}{\min(F_{\text{up}}, \text{UL})} \quad (4)$$

where  $F_{\text{up}} = F + \sigma^+$  corresponds to the flux  $1\sigma$  upper value when there is a detection (with  $F$  the fluxes and  $\sigma^+$  the  $1\sigma$  positive flux error), UL corresponds to the  $3\sigma$  upper limit when there is no detection (as obtained through RapidXMM), and  $F_{\text{low}}$  corresponds to the flux lower value in the case of detection, precisely given by  $F_{\text{low}} = \max(F - \sigma^-, 0)$ , with  $\sigma^-$  as the  $1\sigma$  flux negative



**Fig. 11.** Illustration of the long-term X-ray variability revealed by our method. *Left panel:* distribution of the variability for the multi-instrument sources, including the *XMM-Newton* upper limits. We only show sources consistent with being variable (i.e.,  $\text{Var}_{\text{new}} > 1$ , on the right of the vertical dotted line). The vertical dashed line shows the arbitrary limit for what we consider as significant variability (i.e., pessimistic amplitude above 5). Out of the  $\sim 135\,000$  sources with  $\text{Var}_{\text{new}} > 1$ , only  $\sim 16\,000$  have  $\text{Var}_{\text{new}} > 5$ . *Right panel:* distribution of improvement of variability between all the initial single-catalog sources for which a variability estimate was available, and the final multi-instrument source. The vertical dotted line signifies the limit between single-catalog sources for which the new variability is larger than the prior estimate ( $\sim 49\,000$  sources out of  $\sim 95\,000$ ), and the ones where the new method does not improve the variability estimate ( $\sim 46\,000$ ).

error. Such a definition of  $F_{\text{low}}$  is meant to avoid it being negative number, as this would contaminate the value of  $V$ . If a flux measurement is unconstrained (i.e.,  $F - \sigma^- \leq 0$ ), then this point is essentially ignored in the computation of  $\max(F_{\text{low}})$  if there are other well-constrained data points. Using this definition of the variability  $V$  allows us to estimate simultaneously the amplitude and significance of the variability. If  $V < 1$ , it means that the various data points are consistent at the  $1\sigma$  level, namely, the source appears constant over time. However, if  $V > 1$ , its value gives a lower limit on the actual variability of the underlying physical object. It is important to note here that the variability value we measure is always at best a lower limit of the actual physical variability, due to the sparsity of the X-ray coverage.

Since our cross-matching and upper limits method was meant to improve our constraints on the variability of X-ray objects, we can now assess the effectiveness of our method using this definition of the variability. As was explained in the previous sub-section, our method decreased the number of sources with one data point only, namely, increased the number of sources for which the variability can be estimated. The distribution of variability for the multi-instrument sources is shown in detail in Fig. 11, as well as the gain in variability made using our method. Before the cross-matching, there were 74 030 single-catalog sources with a variability estimate over 1 (out of the 207 966 where the estimate was available, and the 1 258 420 total single-catalog sources), and 4622 with a variability larger than 5. Thanks to our method, out of the resulting 926 753 multi-instrument sources, 618 816 have a variability estimate, which is above 1 for 134 997 multi-catalog sources and above 5 for 15 993 of them. The fraction of variable sources compared to the complete catalog is thus increased from 5% to 15% using our method. The fraction of significantly variable sources ( $V > 5$ ) is also increased from 0.3% to 1.7%. The arithmetic mean gain of variability from the single-catalog sources to the multi-catalog sources is  $\sim 10$  (see Fig. 11), although this is mostly driven by few outlying sources with very large gains. The geometric mean of the variability gain (less contaminated by outliers) is  $\sim 1.4$ . This means that our method is successful in improving the constraint on the X-ray variability of archival sources.

### 3. The STONKS algorithm

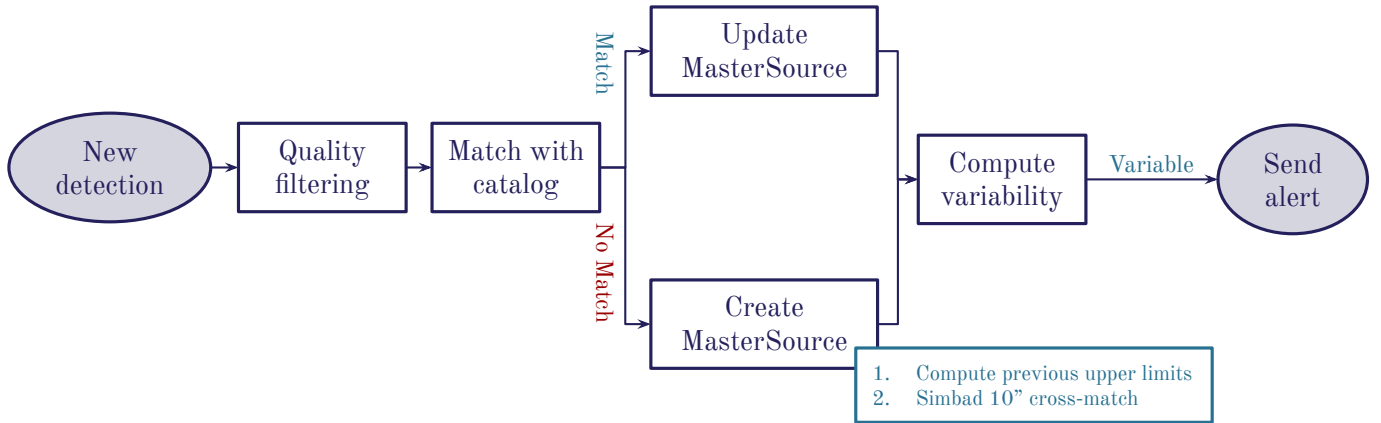
#### 3.1. Motivation and possible implementation within the *XMM-Newton* pipeline

This section presents a possible implementation of our work in the *XMM-Newton* pipeline. This is of course subject to modifications if and when it is to be actually implemented in the data stream.

Currently, the new *XMM-Newton* observations follow a 1-yr proprietary period for non-Heritage data during which the data are only available to the P.I. of the corresponding *XMM-Newton* proposal (see the *XMM-Newton* Announcement of Opportunity<sup>2</sup> for more details). If a transient event was to take place serendipitously within the field of view, and the P.I. failed to detect and identify it, this proprietary period means that any later identification and follow-up processes would take place more than a year after the initial detection. This entails a loss of most of the valuable early-time physical information which could have been gathered if the transient had been immediately detected. For this purpose, we have developed the ‘‘Search for Transient Objects in New detections using Known Sources’’ algorithm (STONKS).

The suggested framework of STONKS is as follows. Once the *XMM-Newton* observational data have been downloaded from the satellite, they go through an automatic processing and source-detection pipeline. As part of the ACDS pipeline, the EPIC summary source list could then be provided to STONKS, in order to check for long-term variability. This would automatically generate a PDF file for each alert in the field of view. This file can be sent to the P.I. of the observation, as part of the PPS products. Additionally, at this point, the pipeline products are checked manually by an *XMM-Newton* scientist (e.g., Watson et al. 2009) – we suggest that the alerts are also checked by the *XMM-Newton* scientist, who will then validate them. After validation, they will be uploaded to a database hosted at IRAP. If the

<sup>2</sup> [https://xmm-tools.cosmos.esa.int/external/xmm\\_user\\_support/documentation/A0policy/Policies\\_Procedures.pdf](https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/A0policy/Policies_Procedures.pdf)



**Fig. 12.** Schematic representation of the workflow of STONKS on a given new *XMM-Newton* detection. The main differences in treatment arise from the result of the cross-match with the archival multi-instrument catalog. A detection is considered “variable” if the associated multi-instrument source (called “MasterSource” here) has a long-term variability larger than five, as defined in 4.

P.I. expressed their agreement and the source is serendipitous, the alerts are then made available on a public web service.

The suggested workflow that would be then followed by each detection is presented in Fig. 12. The new detections would be filtered based on their quality criteria. To be more precise, we require the extent likelihood to be below 6 (to keep only point-like sources), and the detection likelihood to be over 10 ( $\sim 4\sigma$ ) in all EPIC instruments for which the source is in the field of view in order to retain the most reliable sources. Indeed, after initial testing we found that detections for which some instruments had low detection likelihoods but other instruments had sufficient detection likelihood tended to be dominated by spurious detections and instrumental effects. The remaining clean detections would then be first cross-matched with the archival multi-catalog sources, using the  $3\sigma$  position error, and the same ambiguity-solving framework as was used when building the catalog. If the ambiguity cannot be lifted, we cannot safely confirm any long-term variability, so the process stops at this stage. Otherwise, there are two situations: either the source is new and does not match any of the archival sources, in which case the previous possible upper limits would be computed by calling RapidXMM on the source’s position, and a  $10''$  Simbad cross-match performed using the `astroquery` package (Ginsburg et al. 2019). If the source matches the archival catalog without ambiguity (or if this ambiguity is solvable), then the new detection can be added to the multi-catalog source’s history. For both cases, STONKS would then assess the new long-term variability of the source, given this new information. If the multi-catalog source, with the added detection, is overall variable with a pessimistic variability amplitude over five (as was defined in Eq. (4)), a variability alert associated with the detection would be raised.

The output would be presented in the form of a PDF file, with four panels (see examples in Figs. C.2–C.8). The first contains the long-term multi-instrument light curve, including upper limits, extrapolated to the 0.1–12 keV band. The second panel contains the band photometry of each detection, allowing us to assess spectral variability in the source, or spurious flux variability due to extreme softness or hardness of the source (see Sect. 2.2). The third panel contains a  $2' \times 2'$  optical image of the source from the Digital Sky Survey (Lasker et al. 1996), queried using the `astroquery` package. Finally, the fourth panel contains details about the observation itself (observation identifier, date, name of the target), about the detection (identifier of the detection in the observation, position and position error,

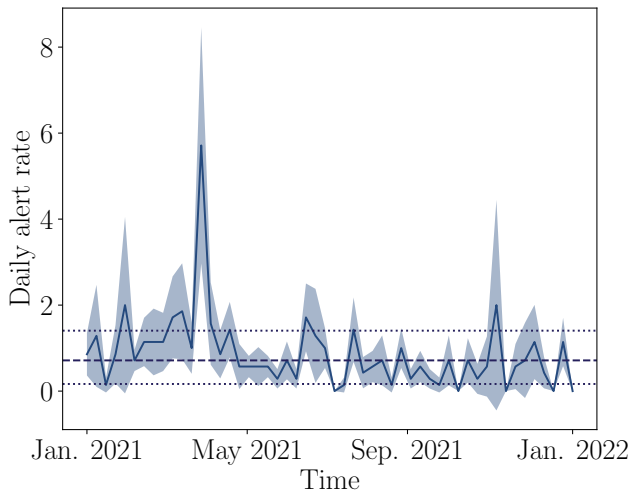
off-axis angle and detection likelihoods in the three EPIC instruments), and about the associated multi-catalog source (type of alert, long-term and short-term variability, and SIMBAD classification if the source was already known). There are four possible types of alerts:

- “High-flux state” if the new detection is the brightest historical state of the multi-catalog source;
- “Low-flux state” if it is the lowest historical state (including lower than past *XMM-Newton* upper limits);
- “First-detection” if this is the first time the source is detected, with prior upper limits being constraining. This is technically similar to “High Flux State”, but might be more sensitive to spurious detections, hence the separate category;
- “Past-variability” in the case where the new detection is between the brightest and dimmest historical states of the multi-catalog source and this source has shown variability in the past.

Finally, we added a warning regarding the spectral assumption. This warning is raised if any of the detections of the source (including the new detection) have a spectral hardness that falls into the 10% hardest or softest detections of its respective catalogs. This could potentially mean that the variability is mis-estimated. The corresponding thresholds are presented in Table B.1. Various examples of serendipitous alerts are available in Appendix C.2. The precise format of the alert PDF file is of course subject to change, depending on the various feedbacks from the *XMM-Newton* scientists and the community, once the service is operational.

We recommend the alert would then be returned to the *XMM-Newton* scientist for manual screening – this would expand the screener’s task, but the expected number of alerts is reasonably low (see Sect. 3.2). Alerts that are not spurious could then be shared using one of the standard community mechanisms. We also intend to upload the alerts as a JSON file to a database hosted at IRAP, that would then be displayed on a publicly available web service (the precise details for this service; for instance: a possible notification system, are yet to be determined). STONKS is currently publicly available through a REST API<sup>3</sup> which takes a *XMM-Newton* EPIC observation source list as an input (POST request) and returns a tarball with all the PDF corresponding the detected variability. The service can be

<sup>3</sup> <https://xcatdb.unistra.fr/stonks/>



**Fig. 13.** Daily alert rate computed on a weekly average. The envelope corresponds to the standard deviation of this daily rate over each week. The dashed and dotted lines correspond to the yearly median and  $1\sigma$  errors on the rate of  $0.7^{+0.7}_{-0.5}$  alerts per day. The large peak at the end of March corresponds to a set of several consecutive observations of Sgr A\*, simultaneous to GRAVITY exposures – the Galactic center is particularly rich in X-ray transient events, either stellar flares or bursts from X-ray binaries.

connected either from a WEB page or through clients such as CURL.

### 3.2. Testing

To assess the validity of our method, we simulated the behavior of the alert system over archival *XMM-Newton* data. We ran STONKS on the 483 observations from 2021 for which the observing mode allows us to observe serendipitous sources, checking variability for 12 584 detections, leading to 315 individual alerts (alert rate of  $\sim 2.5\%$  among all the detections). The various statistics of these alerts are represented in Fig. C.1.

The evolution of the resulting daily alert rate over the testing run can be seen in Fig. 13, with a daily rate of  $0.7^{+0.7}_{-0.5}$  alerts per day. The standard deviation of this daily rate is quite large, as the number of alerts in a given observation is highly dependent on the specific targeted field of view (e.g., the Galactic center is rich in transients).

Out of these 315 alerts, 53 were the target of the observation, while 262 were serendipitous sources. Since the idea behind STONKS is to detect previously unknown transient events, this large fraction ( $\sim 80\%$ ) of serendipitous alerts is encouraging. Even for the target of the observation, an assessment of the long-term variability might be useful for the P.I. of the observation.

Among the 315 alerts, about 40% were linked to past variability events (138), the remaining three categories being about evenly distributed (68 “low-flux state” alerts, 52 “high-flux state” alerts, and 57 “first-detection” alerts). Overall, the target sources have a slightly higher fraction of “past-variability” alerts (28 out of 53) than the serendipitous sources (110 out of 262). This difference is mainly driven by the much larger fraction of “high-flux state” and “first-detection” alerts for serendipitous sources – this is expected for serendipitous transients happening in the field of view. Seven “first-detection” alerts were sent for targets of an observation, showing two limitations of our method. For four of these alerts, they were linked to a high proper motion object (in

this case the M dwarf CN Leo): since our matching methods and upper limit computation work is based on a fixed sky position, high proper motion objects will naturally lead to spurious transient alerts. Correcting this issue would require retrieving the proper motions of the sources in and near the field of view, and compensating it in the various position-dependent steps of our algorithms, which is beyond the scope of our approach. The three remaining alerts were linked to a new TDE detected by eROSITA (eRASSt J045650.3-203750, e.g., Malyali et al. 2023; Liu et al. 2023). While it is reassuring to trigger an alert on a TDE in the field of view, the fact that three alerts were sent out for the same object is due to the fact that STONKS does not update its archival database on new detections. This is meant to avoid spurious detections contaminating the catalog before they are filtered out by manual screening. However, it will lead to multiple alerts being sent out in the case where the source was detected several times since the last data release of the catalogs. This also prevents the detection of variability between two observations of a given data release. This precise approach might be subject to change in ulterior versions of STONKS, with for instance the inclusion of detections from the same data release (after manual screening), with an additional warning about them.

Using the  $10''$  cross-match with Simbad, we retrieve classification for a fraction of the alerts (113 out of 315 – see Fig. C.1). Out of these, 30 correspond to X-ray binaries, 36 to stellar objects, and 47 to galaxies or AGNs. For the remaining alerts, 63 do not have a specific classification in Simbad, which usually indicates that they are part of a large scale catalog (e.g., “X-ray source”, as part of a X-ray telescope catalog with no individual classification). For 139 alerts, they are not at all in Simbad – manual inspection indicates that these are mostly stellar objects. Almost all alerts corresponding to first detections (i.e., using past upper limits) have no Simbad counterpart.

Out of the 315 alerts, the contamination rate is estimated after manual screening to be below 20%. These errors are driven by high proper motion objects, instrumental errors, and more frequently failures of the spectral assumption (as explained in Sect. 2.2). The false alert rate of  $\sim 0.6\%$  presented in Sect. 2.2 can be compared to the  $\sim 2.5\%$  total alert rate per detection we obtained on the 2021 data, confirming the estimated  $\sim 20\%$  contamination. While it is difficult to avoid these issues in our pipeline, the output alert was designed to help manually identify these possibilities. The second panel, showing the band photometry of each X-ray detection, allows us to roughly compare their corresponding spectra and see if they are compatible, despite the flux estimates showing variability. This can be seen for instance in the spurious alert in Fig. C.2: the source being quite hard, the extrapolation between instruments will introduce a bias in the flux estimates, but the spectra are clearly compatible. It is then straight-forward to discard this alert. For the high proper motion objects, the optical view provided in the third panel can allow us to see these objects, as a bright nearby star will appear slightly off-centered from the targeted position. A proper manual screening needs to be performed in order to confidently remove these alerts. Finally, the instrumental errors and spurious detections are hard to exclude in a catalog-oriented approach. Since these alerts will be dealt manually, it will be possible to discard those corresponding to manually flagged spurious detections.

### 3.3. Some variable sources found during the testing run

The idea behind STONKS is to allow us the community to quickly detect X-ray serendipitous transient objects, and follow up on them if relevant. We show in this section a (somewhat

arbitrary) selection of some variable objects found in the 2021 test run of STONKS. These include a possible TDE candidate, AGNs with long-term or short-term (i.e., over the course of a single observation) spectral variability, a flaring star and new XRB and ULX candidates.

For each of these sources, we used the EPIC pn data when available, and the MOS data otherwise. We performed the standard procedure from the *XMM-Newton* data analysis threads<sup>4</sup>, using SAS 19.0.0<sup>5</sup> and Xspec (Arnaud 1996) for the spectral fitting.

### 3.3.1. 4XMM J151509.5+561347: TDE or flaring AGN?

4XMM J151509.5+561347 showed a soft outburst in August 2021 (ObsID 0891801501), with a variability of a factor >13 compared to previous upper limits (see the alert Fig. C.3). Its optical counterpart (SDSS J151509.61+561347.3) is classified as a galaxy (Ahumada et al. 2020), with a photometric redshift of  $0.33 \pm 0.09$ . The nearby galaxy, SDSS J151510.27+561344.7, is brighter and has a spectroscopic redshift of 0.16. Using the photometric redshift of  $0.33 \pm 0.09$ , the peak flux value of  $\sim(7 \pm 1) \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$  translates into a luminosity of  $2.5_{-1.5}^{+2.5} \times 10^{44} \text{ erg s}^{-1}$ . This type of luminosity can be reached by both high accretion episodes in AGN or bright TDEs at their peak. The soft emission is consistent with both as well, however the spectrum (see Fig. C.9) is better explained by an absorbed powerlaw ( $\chi^2/\text{d.o.f.} = 24.5/18$ ,  $\Gamma = 2.7 \pm 0.4$ ) than by an absorbed black body ( $\chi^2/\text{d.o.f.} = 75/18$ ,  $k_B T = 173 \pm 8 \text{ eV}$ ). It is hard to clearly discriminate between these two possibilities based on the spectral shape only. Ideally, a timely X-ray and/or optical follow-up would have allowed us to assess the presence of either AGN or TDE emission, based on the spectral-timing properties of the emission after the peak (e.g., a  $\propto t^{-5/3}$  decay over a few months for a TDE, compared to the red noise expected in an AGN).

### 3.3.2. 4XMM J000532.8+200717: a quasar with variable photon-index

4XMM J000532.8+200717 is a quasar at  $z = 0.119$  (Caccianiga et al. 2008) that showed a significant long-term spectral variability over the 20 yr of available X-ray data (see alert Fig. C.4). It underwent an episode of high emission in the late 2000s, with noticeable *Swift* variability of about an order of magnitude (between luminosities of  $\sim 10^{43}$  to  $\sim 10^{44} \text{ erg s}^{-1}$ ). It is noticeably harder at the peak than in quiescence (see Fig. C.10). The peak spectrum is consistent with an intrinsically absorbed power law ( $N_{\text{H}}^{\text{Peak}} = (1.0 \pm 0.5) \times 10^{20} \text{ cm}^{-2}$  and  $\Gamma^{\text{Peak}} = 3.2 \pm 0.1$ ), with a much softer photon index in the low state and consistent intrinsic absorption ( $N_{\text{H}}^{\text{Low}} = (5 \pm 3) \times 10^{20} \text{ cm}^{-2}$ , and  $\Gamma^{\text{Low}} = 5.2 \pm 0.6$ ). This change is further confirmed by the fact that freezing the photon index at the peak value and fitting only the normalization and absorption on the low state significantly worsens the fit statistics, from  $\chi^2/\text{d.o.f.} = 30/17$  to  $52/18$ .

### 3.3.3. 4XMM J053231.0+170504: a typical stellar flare

4XMM J053231.0+170504 is a star (TYC 1301-1536-1, from Høg et al. 2000) that showed significant X-ray variability by a factor  $\sim 6$  between two *XMM-Newton* observations 2 yr apart

<sup>4</sup> <https://www.cosmos.esa.int/web/xmm-newton/sas-threads>

<sup>5</sup> “Users Guide to the *XMM-Newton* Science Analysis System”, Issue 18.0, 2023 (ESA: XMMNewton SOC).

(see Fig. C.5). Its long-term variability is in fact a consequence of the large short-term flare it underwent during the second *XMM-Newton* observation, which has an impact on the observation-averaged flux (see Fig. C.11). Such X-ray flares, of amplitude  $\sim 5$  and timescale  $\sim 2 \text{ ks}$ , are expected from active stars (e.g., Benz & Güdel 2010).

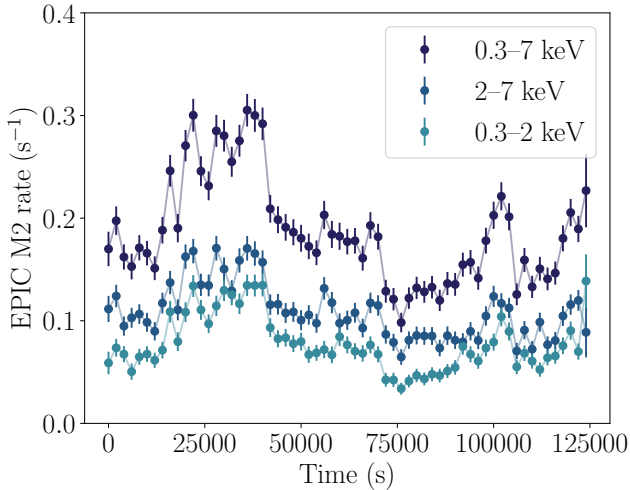
### 3.3.4. 4XMM J000532.8+200717: Quasar with variable photon-index

4XMM J000532.8+200717 is a quasar at  $z = 0.119$  (Caccianiga et al. 2008) that showed a significant long-term spectral variability over the 20 yr of available X-ray data (see alert Fig. C.4). It underwent an episode of high emission in the late 2000s, with noticeable *Swift* variability of about an order of magnitude (between luminosities of  $\sim 10^{43}$  to  $\sim 10^{44} \text{ erg s}^{-1}$ ). It is noticeably harder at the peak than in quiescence (see Fig. C.10). The peak spectrum is consistent with an intrinsically absorbed power law ( $N_{\text{H}}^{\text{Peak}} = (1.0 \pm 0.5) \times 10^{20} \text{ cm}^{-2}$  and  $\Gamma^{\text{Peak}} = 3.2 \pm 0.1$ ), with a much softer photon index in the low state and consistent intrinsic absorption ( $N_{\text{H}}^{\text{Low}} = (5 \pm 3) \times 10^{20} \text{ cm}^{-2}$ , and  $\Gamma^{\text{Low}} = 5.2 \pm 0.6$ ). This change is further confirmed by the fact that freezing the photon index at the peak value and fitting only the normalization and absorption on the low state significantly worsens the fit statistics, from  $\chi^2/\text{d.o.f.} = 30/17$  to  $52/18$ .

### 3.3.5. 4XMM J081909.2+703928: Possibly misclassified ULX candidate

4XMM J081909.2+703928 is a hard X-ray source, appearing in the outskirts of the dwarf galaxy Holmberg II. It showed large variability over the 20 yr of available X-ray data, by a factor of about 300 over short timescales ( $\sim$ days, see alert in Fig. C.6). It is part of the NuSTAR hard X-ray sources catalog (Zappacosta et al. 2018), and an optical spectral follow-up for this study assessed a redshift of  $z=1.27$ , thus making this source an AGN candidate (even blazar candidate, with corresponding variability and lack of spectral change, and peak *Swift* luminosity of  $\sim 10^{46} \text{ erg s}^{-1}$ ). However, the optical counterpart to this source is extremely dim, not even visible in the PanSTARRs survey, meaning that the initial redshift estimate is most likely spurious. The absence of an optical counterpart also excludes the blazar interpretation, which should be bright in optical light as well, seeing as there is no sign of absorption in the X-ray spectrum (see next paragraph).

Ignoring the pre-existing redshift estimate, another possibility is that the source is in the periphery of Holmberg II, and not a background source. This could be strengthened by the presence of a faint UV detection in the *XMM-Newton* Optical Monitor (XMMOM J081909.2+703929, with a UVW1 flux of  $\sim 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ ), without optical counterpart, which could correspond to a faint star cluster. Assuming it is located at the same distance as Holmberg II (i.e., 3.39 Mpc, Karachentsev et al. 2002), the luminosities range from  $10^{37}$  up to  $\sim 3 \times 10^{39} \text{ erg s}^{-1}$ , which is consistent with high-luminosity episodes of an X-ray binary, even reaching ULX-levels of luminosity. The spectrum of a high luminosity episode, for the observation that triggered the alert (ObsID 0864550401) is better fitted by an unabsorbed dual component powerlaw and black body than by a simple unabsorbed powerlaw ( $\chi^2/\text{d.o.f.}$  of 37/31 compared to 65/33), as is shown in Fig. C.12. Such a double component spectrum is characteristic of ULXs and X-ray binaries (e.g., Koliopanos et al. 2017), and less likely for blazars, which are in most cases well-fitted by a single powerlaw component. This



**Fig. 14.** EPIC MOS2 lightcurves of 4XMM J023228.8+202349 (ObsID 0810821801), binned at 2 ks. The soft (0.3–2 keV) and hard (2–7 keV) emission evolve in a similar way, meaning that the change is not due to absorption (which would impact more significantly on the soft emission) but is intrinsic to the powerlaw component. A slight  $\sim 1$ ks lag is visible between the soft and hard emission.

tends to support the idea that this source has been misclassified as a background AGN, and is in fact a possible candidate ULX (or at least X-ray binary) in the outskirts of Holmberg II.

### 3.3.6. 4XMM J013650.6+154011: New candidate XRB

4XMM J013650.6+154011 showed alternating episodes of activity and quiescence over the 20 yr of archival data (see alert Fig. C.7). It displayed variability by a factor  $\sim 10$  on timescales of a few days to a few weeks. This variability was mostly caught by *Swift* and *Chandra*, making any spectral conclusion difficult. Its faint optical counterpart (SDSS J013650.65+154011.3, AB magnitude in the SDSS r band of 20.8), combined with the timescales and amplitude of variability, supports the interpretation of an X-ray binary. This is further confirmed by the peak spectrum, from the observation that triggered the alert (ObsID 0864270101), which is consistent with an absorbed double component emission with a powerlaw and a black body ( $N_{\text{H}} = 6.4_{-3.7}^{+4.5} \times 10^{21} \text{ cm}^{-2}$ ,  $\Gamma = 6.0 \pm 3.0$ ,  $k_{\text{b}}T = 0.66_{-0.13}^{+0.19} \text{ keV}$ ,  $\chi^2/\text{d.o.f.} = 32/32$ , see Fig. C.13), which is typical of X-ray binaries. The other interpretation for such variability would be a blazar, which would have a brighter optical counterpart and is thus excluded.

### 3.3.7. 4XMM J023228.8+202349: Short-term variable absorbed AGN

4XMM J023228.8+202349 is a hard source showing variability by a factor of  $\sim 10$  over timescales of a few days (see alert in Fig. C.8). It is part of the NuSTAR serendipitous catalog as well (Zappacosta et al. 2018), that identified its optical counterpart as a broad-line galaxy at  $z = 0.029$ . The source, in the three available observations, is well fitted with a power law and ionized absorber and a reflection feature (TBabs\*zxipcf\*(zgauss+relxillp)). The brightest *XMM-Newton* observation, which triggered the alert, is short-term variable as well. The EPIC MOS2 lightcurves can be seen in Fig. 14, in several energy bands. There is no difference in the evolution of the soft ( $< 2$  keV) and hard ( $> 2$  keV) bands, meaning

that the change is not in absorption but in the normalization of the power law. The cross-correlation between the soft and hard bands reveals that the soft emission lags slightly ( $\sim 0.8 \pm 0.3$  ks) behind the hard emission (see Fig. C.15). This lag is consistent with the reflection of the hard X-ray corona, which is also confirmed by the spectrum which contains a reflection component (see Fig. C.14). Assuming a constant height of the corona  $h$  for the *relxillp* component, we find that  $h = 5.6 \pm 1.8 r_{\text{g}}$ . The main changes between the observations are the norm of the power law from  $7 \times 10^{-5}$  to  $9 \times 10^{-6}$  and the column density from  $(0.38 \pm 0.12) \times 10^{22} \text{ cm}^{-2}$  to  $(6.5 \pm 2.5) \times 10^{22} \text{ cm}^{-2}$ . The lag of  $\sim 0.8 \pm 0.3$  ks is indicative of a size of  $(2.4 \pm 0.9) \times 10^{11} \text{ m}$ . Assuming this size is the corona-disk distance, namely,  $\sim h$ , we find  $r_{\text{g}} \approx 0.4_{-0.2}^{+0.4} \times 10^{11} \text{ m}$ , namely,  $M_{\text{BH}} \approx 2.7_{-1.3}^{+2.7} \times 10^7 M_{\odot}$ .

## 4. Discussion

### 4.1. Implementation in the XMM-Newton pipeline and alert dissemination

STONKS is designed to automatically search for transients in *XMM-Newton* EPIC data and can be used to find them in quasi-real time if run at the time of the automatic pipeline processing. These alerts can then be shared with the P.I. of the observation, and with the community, in order to ensure that no transient is overlooked. In essence, it is the *XMM-Newton* equivalent to the newly implemented transient detector for the *Swift* pipeline (Evans et al. 2023). Another future possibility for making use of these alerts is to create synergies with data brokers for the *Vera C. Rubin* Observatory, such as Fink (Möller et al. 2021).

### 4.2. Main limitations and expected results

As explained in Sect. 3.2, the main limitations of our method are (in decreasing order based on the contamination rate) the failure of the spectral extrapolation assumption in the case of very hard or very soft sources, the presence of instrumental errors and spurious detections, and high proper motion objects for which astrometry-based matching is not straight forward. These issues can be mitigated by manual screening of the produced alert files. Our bayesian cross-match method was successful in avoiding spurious variability based on wrong associations, as no such alert was triggered in the 2021 test run.

The alert rate obtained from the 2021 test run is expected to be representative of the general rate of alerts raised for transients with a variability of at least a factor 5 detected with *XMM-Newton*. While these variable objects are dominated by usual AGN variability and stellar flares, a number of more exotic sources have already been detected in the test run. Only serendipitously detected sources were presented in Sect. 3.2, as the philosophy behind STONKS is to detect serendipitous variable objects. However, STONKS also would have raised alerts for some variable targeted objects, among which are two TDE candidates – eRASS1 J045650.3-203750, and 4XMM J011104.7-455843. The fact that STONKS was able to catch these targeted objects means that we would also have caught them if they had been serendipitous detections, confirming the efficiency of STONKS.

### 4.3. Updating the archival catalog

At the time of publication of this work, some catalogs that have been used are already slightly outdated (for instance for *XMM-Newton* by two data releases). However, it is our intention to



update regularly the archival catalog in use, in order to better be able to detect new transient events. In particular, the inclusion of the eFEDS catalog was meant as a proof-of-concept that, once the eROSITA data from the first all-sky survey are released, it will easily be taken into consideration for future detections. It should theoretically provide us systematically with one data point for comparison, for each new *XMM-Newton* detection – or a possibly constraining eROSITA prior upper-limits in the case of an absence of match between the catalogs. The similarity between the *XMM-Newton* and eFEDS sources in terms of flux (see Fig. D.1) is reassuring for the future transient alerts. Additionally, the upcoming *Chandra* and *XMM-Newton* slew data releases of all observations after 2014, as well as regularly updated versions of the Living *Swift*-XRT Point Sources catalog (LSXPS, Evans et al. 2023), will also be taken into account.

#### 4.4. Data mining the archival catalog

While the focus of this work has been on quasi-real time transient detection, the archival catalog that was built as a by-product of our method is a goldmine for archival variability study. During its elaboration, we have used several criteria to mine it, looking for specific sources of interest. In particular, it allowed us to find a new transient ultra-luminous X-ray source in NGC 7793 with a candidate pulsation (Quintin et al. 2021), and a new candidate source of quasi-periodic eruptions in an optically-detected TDE (Quintin et al. 2023). Other long-term variable sources, such as new X-ray TDE candidates, have been found in this archival catalog (Quintin et al., in prep.).

Our work has mostly focused on long-term X-ray variability estimation and detection. However, others may make use of this archival multi-instrument X-ray catalog for other purposes. For this reason, the cross-matched catalog is made available on both Zenodo<sup>6</sup> and the CDS. These files will be updated with each new version of the archival catalog.

## 5. Conclusions

In this paper, we present a new quasi-real time transient detection system for the *XMM-Newton* pipeline, STONKS. The general idea of STONKS is to automatically compare any new *XMM-Newton* detection to our archival knowledge of the X-ray sky at this position in order to assess the long-term variability of the underlying X-ray source.

It required a first step of collating most available archival X-ray data. We used the *XMM-Newton* pointed, slew, and stacked catalogs, the *Chandra* catalog, the *Swift* point-like sources catalog, the ROSAT survey and pointed catalogs, and finally the eROSITA eFEDS catalog. We used relatively stringent quality criteria in order to avoid spurious detections, and in particular only kept point-like sources. The catalogs were then matched together two by two at first, with ambiguous correlations being dealt with using a Bayesian framework similar to that of NWAY (Salvato et al. 2018). The main difference between our method and NWAY is that, at the two-by-two matching phase, catalogs are considered in a symmetrical way (whereas NWAY is inherently asymmetrical, looking for the best counterpart for each source of the primary catalog in the secondary catalog). The two-by-two correlations are then merged in a multi-instrument catalog, in a conservative manner by refusing any “chain-like” association between more than two catalogs. This provided us with a catalog of 926 753 multi-instrument sources, with 15% of them

containing information from multiple catalogs. In order to be able to compare flux values between instruments with varying energy bandwidth, we need to convert these instrument-specific fluxes to a common band and, more precisely, the largest possible band using these catalogs, 0.1–12 keV. This extrapolation is done using a fix spectral assumption (absorbed power law with  $N_{\text{H}} = 3 \times 10^{20} \text{ cm}^{-2}$  and  $\Gamma = 1.7$ ). This assumption is reasonable for most X-ray sources, and is used in the *XMM-Newton* catalogs. We estimated the rate of false positives to be about 0.5% of the total detections and less than ~20% of the alerts, corresponding to the spectrally hardest and softest sources. We then called RapidXMM on the position of all the sources lying in the 4XMM DR11 footprint, in order to retrieve *XMM-Newton* EPIC 0.2–12 keV flux  $3\sigma$  upper limits (even in the case of non-*XMM-Newton* sources, for instance very faint *Chandra* sources, or hopefully transient events). This provided us with 2.8 million flux upper limits, out of which ~200 000 are constraining (i.e., lower than the minimum multi-instrument flux).

Once this archival X-ray multi-instrument catalog was built and *XMM-Newton* upper limits computed, we developed the STONKS pipeline, which takes new detections from an *XMM-Newton* observation and compares them to this catalog. The variability is defined as the pessimistic ratio between the maximum and minimum 0.1–12 keV fluxes of the source (pessimistic in the sense that the error bars are subtracted for the maximum and added for the minimum). If it is above five, a variability alert figure is built, with the long-term multi-instrument light curves, spectra (using catalog-specific band photometry), a  $2' \times 2'$  optical view, and a summary about the source’s properties. We tested the behavior of STONKS on 483 *XMM-Newton* observations from 2021. A daily alert rate of  $0.7^{+0.7}_{-0.5}$  alerts per day was obtained, with 80% of the sources being serendipitous and 40% not in Simbad, which is encouraging for the prospect of finding new transient events. Some of the sources of interest were analysed, including a candidate TDE, a quasar with variable spectrum, a new candidate ULX and a new candidate X-ray binary, a hard flare from an AGN and, finally, a variable AGN showing ionized absorption and a reflection component in the high state. Two confirmed TDEs that were targets of their observation were also detected, further confirming the ability of STONKS to find these variable objects. After manual screening, we estimated the false alarm rate to be below 20%, mostly due to failures of the spectral assumption (i.e., the source is spectrally too hard or soft). We have specifically designed the alert figure to allow us to easily visually identify this situation as well as automatically raising a warning, using the catalog-specific band photometry. The STONKS alerts should be manually screened to ensure their quality.

STONKS could provide the X-ray community with a new ability to detect and follow up on astrophysical transients, and possibly build synergies with future multi-wavelength transients, such as the *Vera C. Rubin* Observatory for instance. This could be very useful with respect to furthering our understanding of many astrophysical transient events. The archival multi-instrument catalog is a by-product of our method, but it can have many uses on its own. It has been made available to the community and will be kept up to date with ulterior data releases, including the first eROSITA sky surveys.

*Acknowledgements.* Softwares: *numpy* (Harris et al. 2020), *matplotlib* (Hunter 2007), *astropy* (Astropy Collaboration et al. 2013, 2018, 2022), *astroquery* (Ginsburg et al. 2019), *CMasher* (van der Velden 2020), *Xspec* (Arnaud 1996), *SAS* (Gabriel et al. 2004). This research has made use of *hips2fits*, a tool developed at CDS, Strasbourg, France aiming at extracting FITS images from

<sup>6</sup> <https://zenodo.org/doi/10.5281/zenodo.10634292>

HiPS sky maps with respect to a WCS. The authors thank the anonymous referee for useful comments that helped improve the quality of this paper. Some of this work was done as part of the XMM2ATHENA project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101004168, the XMM2ATHENA project. E.Q. thanks Mickaël Coriat for his help on the web service implementation. E.Q., N.A.W., and E.K. acknowledge the CNES who also supported this work. I.T. gratefully acknowledges the support by Deutsches Zentrum für Luft- und Raumfahrt (DLR) through grants 50 OX 1901 and 50 OX 2301.

## References

- Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, *ApJS*, **249**, 3
- Arcodia, R., Merloni, A., Nandra, K., et al. 2021, *Nature*, **592**, 704
- Arcodia, R., Miniutti, G., Ponti, G., et al. 2022, *A&A*, **662**, A49
- Arnaud, K. A. 1996, in *Astronomical Data Analysis Software and Systems V*, 101, 17
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, *ApJ*, **935**, 167
- Atteia, J.-L., Cordier, B., & Wei, J. 2022, *Int. J. Mod. Phys. D*, **31**, 2230008
- Bachetti, M., Harrison, F. A., Walton, D. J., et al. 2014, *Nature*, **514**, 202
- Barret, D., & Dupourqué, S. 2024, *A&A*, **686**, A133
- Bellm, E. 2014, *The Third Hot-wiring the Transient Universe Workshop (HTU-III)*, eds. P.R. Wozniak, et al., <http://www.slac.stanford.edu/econf/C131113.1/>, 27
- Benz, A. O., & Güdel, M. 2010, *ARA&A*, **48**, 241
- Boch, T. 2019, in *Astronomical Data Analysis Software and Systems XXXVI*, 521, 487
- Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, *A&A*, **588**, A103
- Boller, T., Schmitt, J. H. M. M., Buchner, J., et al. 2022, *A&A*, **661**, A8
- Budavári, T., & Szalay, A. S. 2008, *ApJ*, **679**, 301
- Caccianiga, A., Severgnini, P., Della Ceca, R., et al. 2008, *A&A*, **477**, 735
- Chakraborty, J., Kara, E., Masterson, M., et al. 2021, *ApJ*, **921**, L40
- Chen, X., Qiu, Y., Li, S., & Liu, F. K. 2022, *ApJ*, **930**, 122
- Coughlin, E. R., & Nixon, C. J. 2019, *ApJ*, **883**, L17
- Evans, I. N., Primini, F. A., Miller, J. B., et al. 2020a, *Am. Astron. Soc. Meet. Abstr.*, **235**, 154.05
- Evans, P. A., Page, K. L., Osborne, J. P., et al. 2020b, *ApJS*, **247**, 54
- Evans, P. A., Page, K. L., Beardmore, A. P., et al. 2023, *MNRAS*, **518**, 174
- Fernique, P., Boch, T., Donaldson, T., et al. 2014, *IVOA Recommendation 02 June 2014*, 602
- Franchini, A., Bonetti, M., Lupi, A., et al. 2023, *A&A*, **675**, A100
- Freund, S., Czesla, S., Robrade, J., Schneider, P. C., & Schmitt, J. H. M. M. 2022, *A&A*, **664**, A105
- Gabriel, C., Denby, M., Fyfe, D. J., et al. 2004, in *Astronomical Data Analysis Software and Systems (ADASS) XIII*, 314, 759
- Gezari, S. 2021, *ARA&A*, **59**, 21
- Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, *AJ*, **157**, 98
- Giustini, M., Miniutti, G., & Saxton, R. D. 2020, *A&A*, **636**, A2
- Graham, M. J., Ross, N. P., Stern, D., et al. 2020, *MNRAS*, **491**, 4925
- Gürpide, A., Godet, O., Koliopanos, F., Webb, N., & Olive, J.-F. 2021, *A&A*, **649**, A104
- Hammerstein, E., Velzen, S. v., Gezari, S., et al. 2022, *ApJ*, **942**, 9
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, *A&A*, **355**, L27
- Israel, G. L., Belfiore, A., Stella, L., et al. 2017, *Science*, **355**, 817
- Ivezić, Z., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Jonker, P. G., Glennie, A., Heida, M., et al. 2013, *ApJ*, **779**, 14
- Kaaret, P., Feng, H., & Roberts, T. P. 2017, *ARA&A*, **55**, 303
- Karachentsev, I. D., Dolphin, A. E., Geisler, D., et al. 2002, *A&A*, **383**, 125
- Kaur, K., Stone, N. C., & Gilbaum, S. 2023, *MNRAS*, **524**, 1269
- King, A. 2020, *MNRAS*, **493**, L120
- King, A. 2022, *MNRAS*, **515**, 4344
- Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, *PASP*, **129**, 104502
- Koliopanos, F., Vasilopoulos, G., Godet, O., et al. 2017, *A&A*, **608**, A47
- König, O., Saxton, R. D., Kretschmar, P., et al. 2022a, *Astron. Comput.*, **38**, 100529
- König, O., Wilms, J., Arcodia, R., et al. 2022b, *Nature*, **605**, 248
- Lasker, B. M., Doggett, J., McLean, B., et al. 1996, in *Astronomical Data Analysis Software and Systems V*, 101, 88
- Li, D., Starling, R. L. C., Saxton, R. D., Pan, H.-W., & Yuan, W. 2022, *MNRAS*, **512**, 3858
- Linial, I., & Metzger, B. D. 2023, *ApJ*, **957**, 34
- Liu, Z., Malyali, A., Krumpel, M., et al. 2023, *A&A*, **669**, A75
- Luca, A. D., Salvaterra, R., Belfiore, A., et al. 2021, *A&A*, **650**, A167
- Madsen, K. K., Beardmore, A. P., Forster, K., et al. 2017, *AJ*, **153**, 2
- Malyali, A., Liu, Z., Merloni, A., et al. 2023, *MNRAS*, **520**, 4209
- Margutti, R., Metzger, B. D., Chornock, R., et al. 2019, *ApJ*, **872**, 18
- Miniutti, G., Saxton, R. D., Giustini, M., et al. 2019, *Nature*, **573**, 381
- Miniutti, G., Giustini, M., Arcodia, R., et al. 2023a, *A&A*, **674**, L1
- Miniutti, G., Giustini, M., Arcodia, R., et al. 2023b, *A&A*, **670**, A93
- Möller, A., Peloton, J., Ishida, E. E. O., et al. 2021, *MNRAS*, **501**, 3272
- Pallavicini, R., Golub, L., Rosner, R., et al. 1981, *ApJ*, **248**, 279
- Pan, X., Li, S.-L., Cao, X., Miniutti, G., & Gu, M. 2022, *ApJ*, **928**, L18
- Petroff, E., Hessels, J. W. T., & Lorimer, D. R. 2019, *A&AR*, **27**, 4
- Pineau, F.-X., Motch, C., Carrera, F., et al. 2011, *A&A*, **527**, A126
- Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, *A&A*, **647**, A1
- Preibisch, T., Kim, Y.-C., Favata, F., et al. 2005, *ApJS*, **160**, 401
- Pye, J. P., Rosen, S., Fyfe, D., & Schröder, A. C. 2015, *A&A*, **581**, A28
- Quintin, E., Webb, N. A., Gürpide, A., Bachetti, M., & Fürst, F. 2021, *MNRAS*, **503**, 5485
- Quintin, E., Webb, N. A., Guillot, S., et al. 2023, *A&A*, **675**, A152
- Rees, M. J. 1988, *Nature*, **333**, 523
- Remillard, R. A., & McClintock, J. E. 2006, *ARA&A*, **44**, 49
- Ruiz, A., Georgakakis, A., Gerakakis, S., et al. 2022, *MNRAS*, **511**, 4265
- Salvato, M., Buchner, J., Budavári, T., et al. 2018, *MNRAS*, **473**, 4937
- Salvato, M., Wolf, J., Dwelly, T., et al. 2022, *A&A*, **661**, A3
- Saxton, R. D., Read, A. M., Esquej, P., et al. 2008, *A&A*, **480**, 611
- Saxton, R., Read, A., Esquej, P., Miniutti, G., & Alvarez, E. 2011, arXiv e-prints [arXiv:1106.3507]
- Saxton, C. J., Perets, H. B., & Baskin, A. 2018, *MNRAS*, **474**, 3307
- Saxton, R., Komossa, S., Auchettl, K., & Jonker, P. G. 2021, *Space Sci. Rev.*, **217**, 18
- Saxton, R. D., König, O., Descalzo, M., et al. 2022, *Astron. Comput.*, **38**, 100531
- Shu, X. W., Wang, S. S., Dou, L. M., et al. 2018, *ApJ*, **857**, L16
- Smith, M. J. S. 2022, <https://xmmweb.esac.esa.int/docs/documents/CAL-TN-0018.pdf>
- Sniegowska, M., Czerny, B., Bon, E., & Bon, N. 2020, *A&A*, **641**, A167
- Sniegowska, M., Grzedziński, M., Czerny, B., & Janiuk, A. 2023, *A&A*, **672**, A19
- Song, X., Walton, D. J., Lansbury, G. B., et al. 2020, *MNRAS*, **491**, 1260
- Stelzer, B., Marino, A., Micela, G., López-Santiago, J., & Liefke, C. 2013, *MNRAS*, **431**, 2063
- Sun, L., Shu, X., & Wang, T. 2013, *ApJ*, **768**, 167
- Taylor, M. B. 2006, in *Astronomical Data Analysis Software and Systems XV*, 351, 666
- Tranin, H., Godet, O., Webb, N., & Primorac, D. 2022, *A&A*, **657**, A138
- Traulsen, I., Schwöpe, A. D., Lamer, G., et al. 2019, *A&A*, **624**, A77
- Vagnetti, F., Turriziani, S., & Trevese, D. 2011, *A&A*, **536**, A84
- van der Velden, E. 2020, *J. Open Source Softw.*, **5**, 2004
- Wang, M., Yin, J., Ma, Y., & Wu, Q. 2022a, *ApJ*, **933**, 225
- Wang, Y., Jiang, N., Wang, T., et al. 2022b, *ApJ*, **930**, L4
- Watson, M. G., Schröder, A. C., Fyfe, D., et al. 2009, *A&A*, **493**, 339
- Webb, N. A., Schwöpe, A., Zolotukhin, I., Lin, D., & Rosen, S. R. 2018, *A&A*, **615**, A133
- Webb, N. A., Coriat, M., Traulsen, I., et al. 2020, *A&A*, **641**, A136
- Webb, N. A., Carrera, F. J., Schwöpe, A., et al. 2023, *Astron. Nachr.*, **344**, e220102
- Wevers, T., Pasham, D. R., Jalan, P., Rakshit, S., & Arcodia, R. 2022, *A&A*, **659**, L2
- White, N. E., Giommi, P., & Angelini, L. 1994, *Int. Astron. Union Circ.*, **6100**, 1
- White, R. J., & Peterson, B. M. 1994, *PASP*, **106**, 879
- Wu, S., Coughlin, E. R., & Nixon, C. 2018, *MNRAS*, **478**, 3016
- Xian, J., Zhang, F., Dou, L., He, J., & Shu, X. 2021, *ApJ*, **921**, L32
- Zappacosta, L., Comastri, A., Civano, F., et al. 2018, *ApJ*, **854**, 33
- Zhao, Z. Y., Wang, Y. Y., Zou, Y. C., Wang, F. Y., & Dai, Z. G. 2022, *A&A*, **661**, A55

## Appendix A: Summary of the catalog quality filters

### – *XMM-Newton* Pointed:

1. "EP\_8\_DET\_ML>8 & SUM\_FLAG<3", which are the detection likelihood and summary of quality flags, selected to ensure a good quality detection. The condition on the likelihood means that these detections are at a  $\sim 5\sigma$  level. The SUM\_FLAG is a summary of the various detection quality flags (see details on the *XMM-Newton* catalog website<sup>7</sup>, and this value means that the detection was cleared by the manual screening;
2. "EP\_EXTENT==0" (which actually means that EP\_EXTENT <6"), allowing us to exclude detections where the source is extended

### – *XMM-Newton* Stacked:

1. "EP\_DET\_ML>8 & STACK\_FLAG<3 & EXTENT==0", same criteria as for the detection catalog;
2. "TAUNAME\_4XMMDR11==Null", i.e., we only keep detections which are new and not present in the detection 4XMM DR11 catalog

### – *XMM-Newton* Slew:

1. We used the Clean catalog, so "DET\_ML>10.5" (i.e.,  $\sim 6\sigma$  detection) and some sources have been manually excluded;
2. "EXT\_B8==0 | (NULL\_EXT\_ML\_B8 & (EXT\_B6==0. | EXT\_B7==0.))": we only keep point-like sources, but as a detection can happen in any of the bands 8, 6 or 7, we only ask for the extent to be zero in the actual detection band.

### – *Chandra*:

1. "likelihood\_class==TRUE": select only good quality detections. This is based on a detection likelihood threshold that is variable and function of the observation's properties, and is computed to have at most 0.1 false detection per stacks of observations. Such good quality detections are the ones above this threshold;
2. "name" does not end with "X" & "conf\_code<256": removes the extended sources, and the sources that lie within the extension of another source;
3. "conf\_flag==FALSE": removes detections for which the association with a single *Chandra* source is ambiguous, due to off-axis PSF;
4. Detections and upper limits are separated based on the filter "flux\_aper\_b==0."

### – *Swift*:

1. We used the clean source sub-sample, so the detection quality flag is 0 or 1, the field quality flag is 0 or 1, and only datasets of quality flag 0 or 1 are used.
2. This catalog natively only contains sources seen as point-like for *Swift*;
3. We excluded the detections where "Rate0==0.0"; while these might naively correspond to upper limits, the 2SXPS catalog is not supposed to contain such upper limits. These  $\sim 1000$  detections are thus considered as spurious, and removed;

### – ROSAT Survey:

1. "EXT\_ML>8 & S\_flag==0", to only keep good quality detections
  2. "EXT==0.", to only keep point-like sources. We also removed any source closer than 10' to a *XMM-Newton* or *Chandra* bright extended source, as some point-like sources for ROSAT are extended for better spatially resolved instruments, meaning that the source is excluded and ulterior associations are spurious;
- ### – ROSAT Pointed:
1. "Qflag>=8", which is a summary flag allowing us to exclude any extended source, located-within-extension source, or any type of spurious detection;
  2. We also removed any source closer than 10' to a *XMM-Newton* or *Chandra* bright extended source;
- ### – eROSITA:
1. "DET\_LIKE>8" (i.e.,  $\sim 5\sigma$  detection), to keep only good quality detections;
  2. "EXT==0.", exclude extended sources

<sup>7</sup> [http://xmmssc.irap.omp.eu/Catalogue/4XMM-DR11/col\\_flags.html](http://xmmssc.irap.omp.eu/Catalogue/4XMM-DR11/col_flags.html)

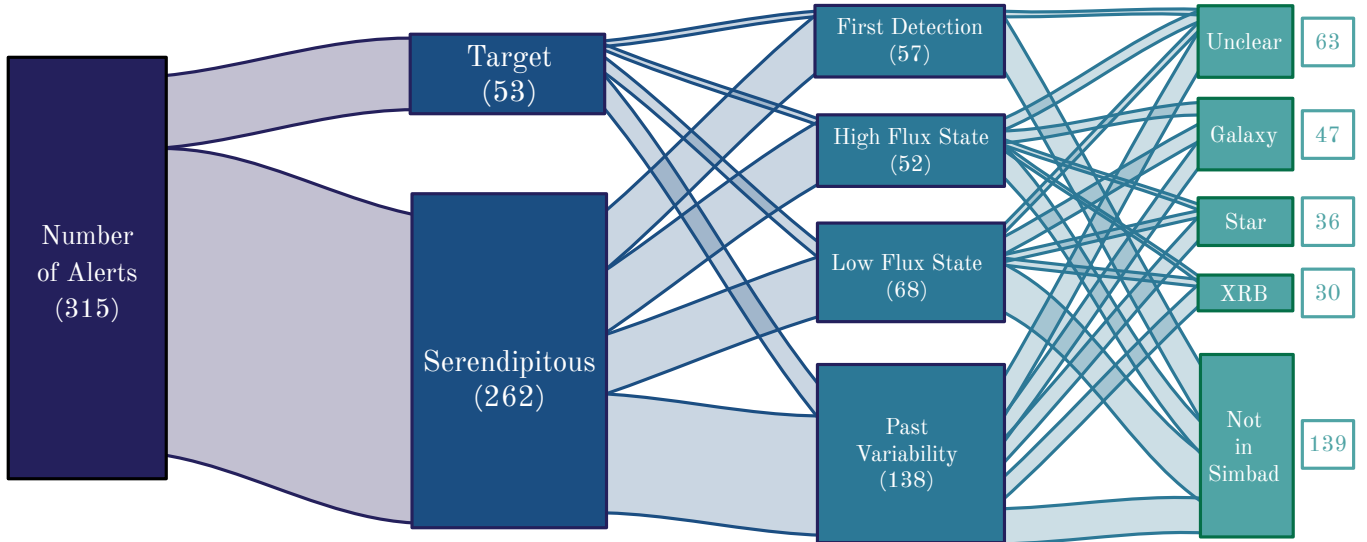
## Appendix B: Energy conversion factors

Catalog	Total Band	Total fraction	Soft band	Soft threshold	Hard band	Hard threshold
XMM-DR11, DR11s, SL2	0.2–12 keV	0.999	0.2–2 keV	<-0.42	2–12 keV	>0.88
2SXPS	0.3–10 keV	0.9	0.3–2 keV	<-0.4	2–10 keV	>0.84
CSC2	0.5–7 keV	0.69	0.5–2 keV	<-0.33	2–7 keV	>0.774
eFEDS	0.2–4.5 keV	0.60	0.2–2 keV	<-0.62	2–4.5 keV	>0.45
RASS, WGACAT	0.1–2.4 keV	0.35	0.2–2.4 keV	N/A	N/A	N/A

**Table B.1.** The various total, soft and hard energy bands of the catalogs considered in this work. For the total band, we indicate the fraction of reference total flux (0.1–12 keV for a spectrum with  $\Gamma = 1.7$  and  $N_H = 3 \times 10^{20} \text{ cm}^{-2}$ ) this band contains. This allows us to calibrate the various catalogs, assuming this underlying spectral shape. For the soft and hard bands, we show the threshold in hardness ratio above (resp. below) which a detection is in the 10% hardest (resp. softest) of its catalogs, which could lead to errors of factor of  $\sim 2$  in the flux calibration and, thus, in the variability computation.

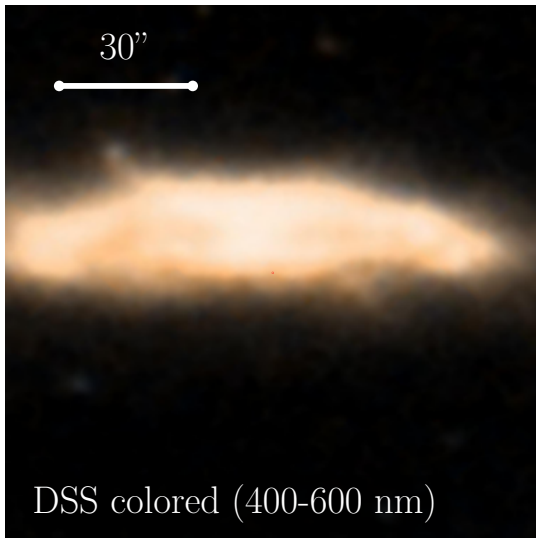
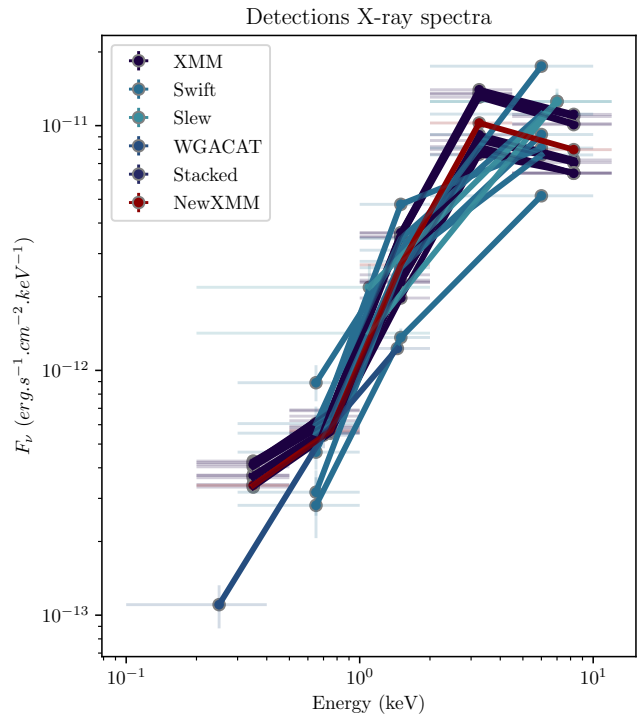
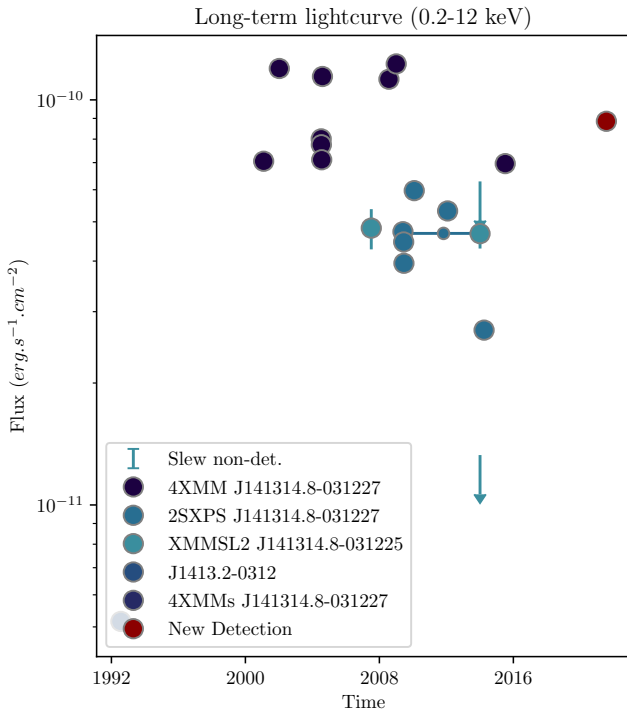
## Appendix C: STONKS alert

### Appendix C.1. Statistics



**Fig. C.1.** Statistics of the test run of STONKS on a part of the 2021 *XMM-Newton* archive. The height of the boxes and branches are proportional to the number of alerts – we have chosen to not display the exact numbers for the sake of readability. The main takeaways are the high fraction of serendipitous alerts and the high fraction of sources that are either 1) not in Simbad or 2) in Simbad, but with no details on the nature of the object. This shows the potential of STONKS to uncover new hidden transients.

Appendix C.2. Alerts from sources of interest from the 2021 STONKS test run



ObsID ..... 0880550401  
 Date Obs ..... 2021-07-22T09:59:21  
 Target Name ..... NGC 5506  
 Exposure Time ..... 56946 s

SRCNUM ..... 1  
 Source RA ..... 213.31 / 14:13:15  
 Source Dec ..... -3.21 / -03:12:27  
 Position Error ..... 0.01"  
 Off-axis Angles ..... PN: 1.2', M1: 0.2', M2: 1.2'  
 Instruments DetML .....  
 PN: 1860208.9, M1: nan, M2: nan, EP: 1860208.9

Type of Alert ..... Past Variability  
 Long-term Variability ..... 14.1  
 Short-term Variability ..... True  
 Simbad ..... Unknown ([MRN2004] NGC 5506 B2)

!!! Warning: extreme spectrum might impact variability !!!

Fig. C.2. Example of spurious variability due to the hardness of the source (here, due to the amount of absorption in the host galaxy). The tiny red dot in the middle of the DSS image (bottom left) is the  $1\sigma$  positional error circle of the X-ray source.

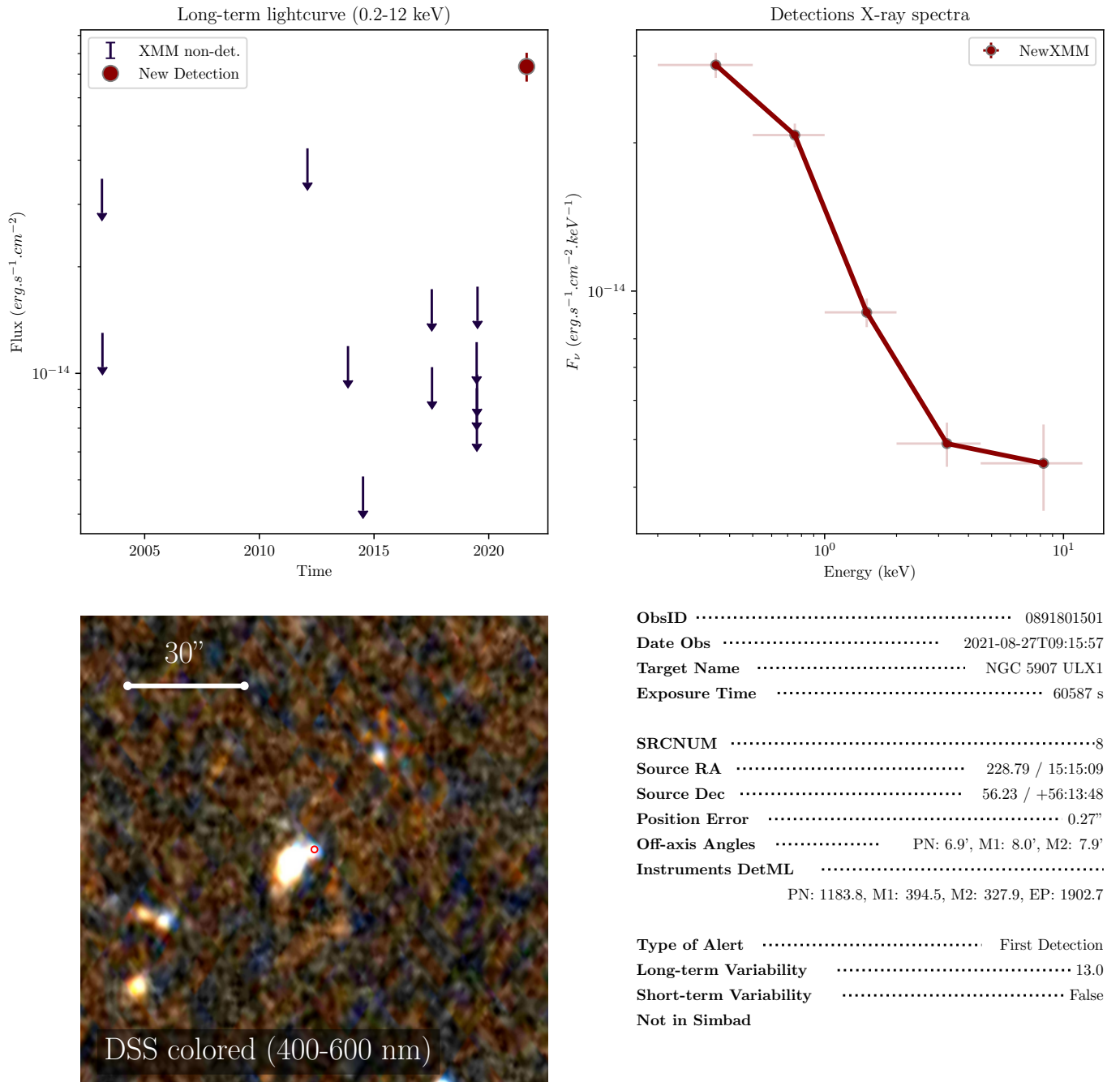
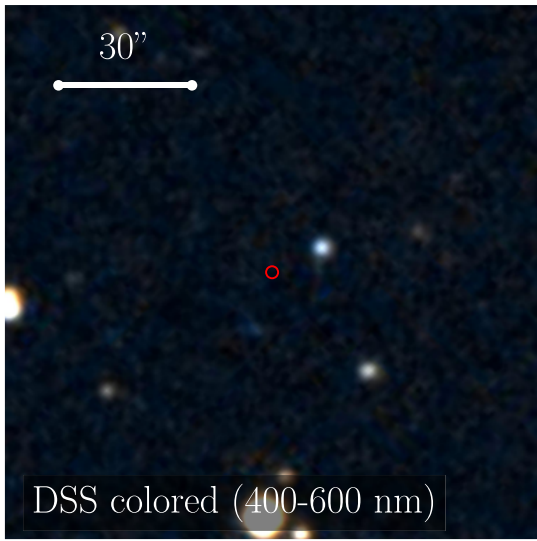
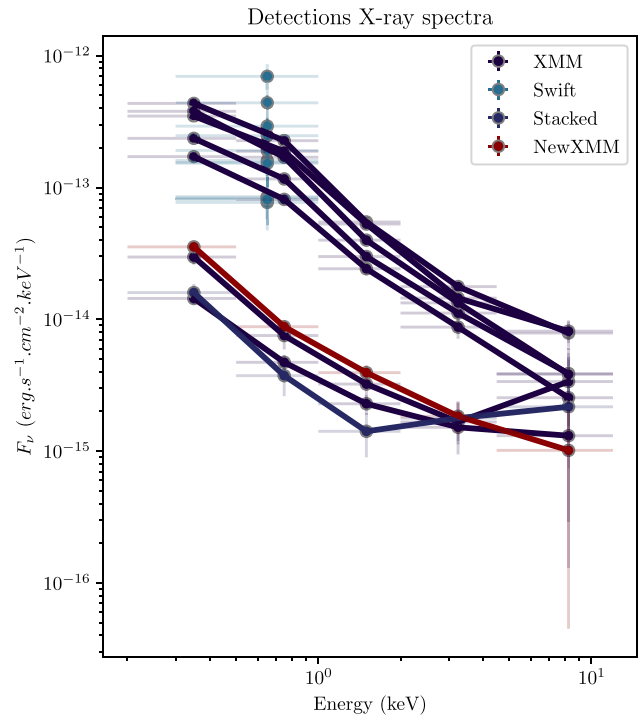
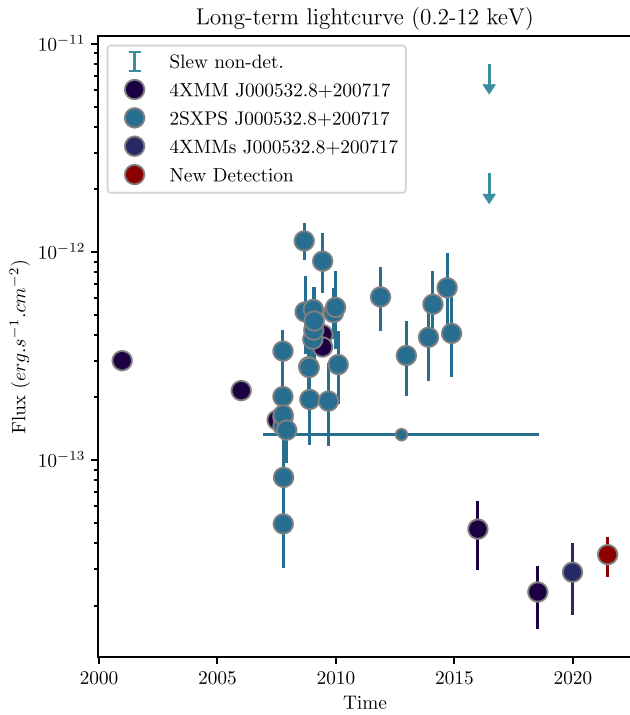


Fig. C.3. Example of an alert sent out by STONKS: a possible TDE candidate or a flaring AGN.



ObsID ..... 0842761301  
 Date Obs ..... 2021-06-22T06:26:37  
 Target Name ..... Mrk 335  
 Exposure Time ..... 89364 s

SRCNUM ..... 16  
 Source RA ..... 1.39 / 00:05:32  
 Source Dec ..... 20.12 / +20:07:17  
 Position Error ..... 0.45"  
 Off-axis Angles ..... PN: 13.1', M1: 12.1', M2: 11.7'  
 Instruments DetML .....  
 PN: 428.7, M1: 149.3, M2: 100.9, EP: 675.3

Type of Alert ..... Past Variability  
 Long-term Variability ..... 21.5  
 Short-term Variability ..... True  
 Simbad ..... Galaxy (2MASS J00053278+2007176)

!!! Warning: extreme spectrum might impact variability !!!

Fig. C.4. Example of an alert sent out by STONKS: a quasar with variable photon-index.

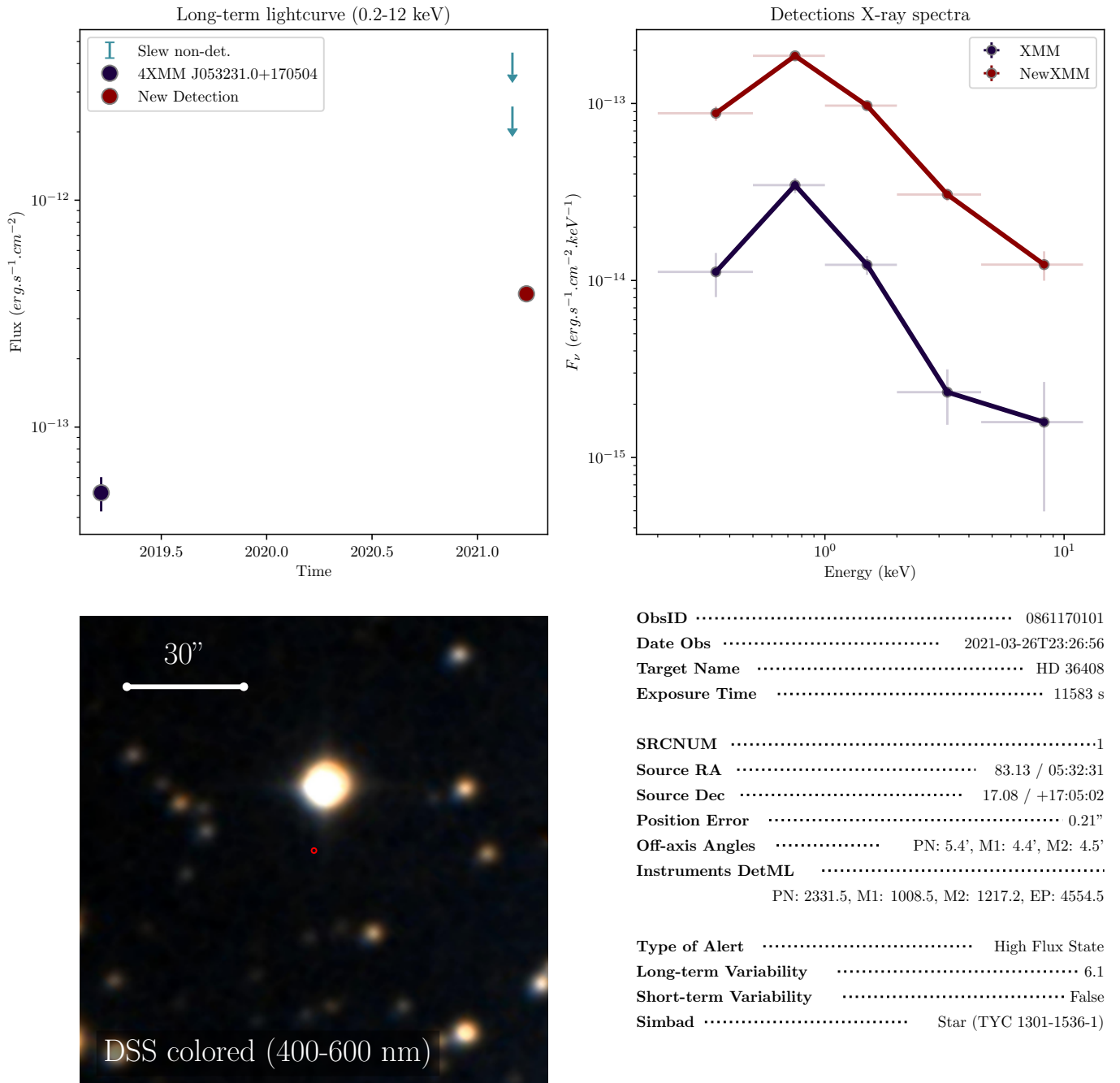
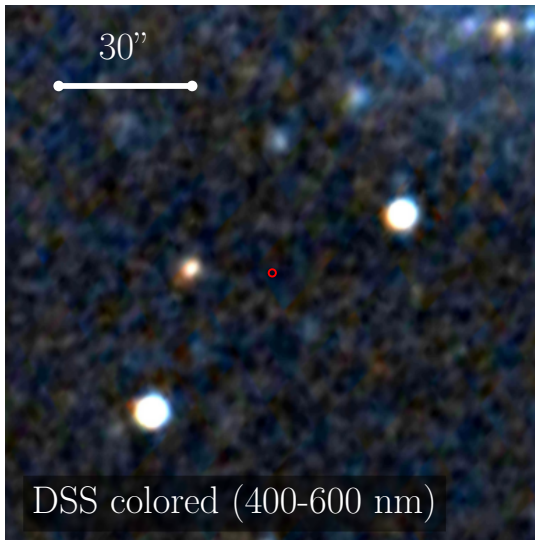
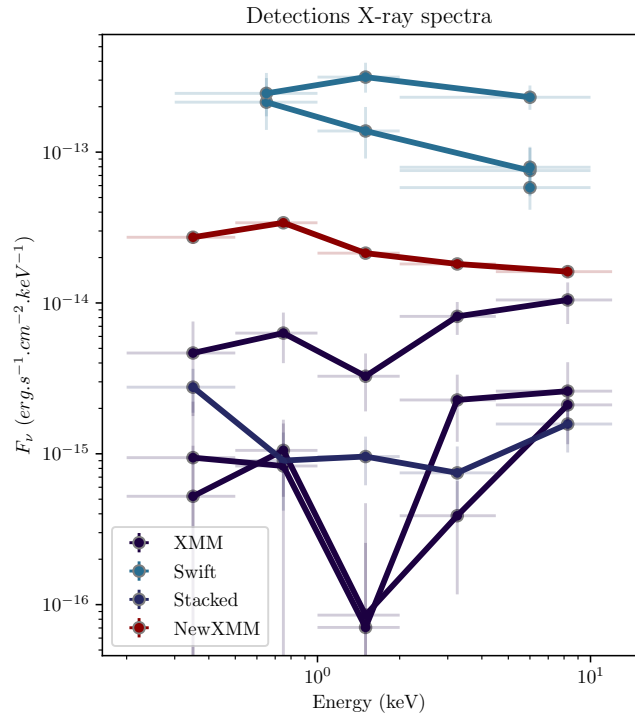
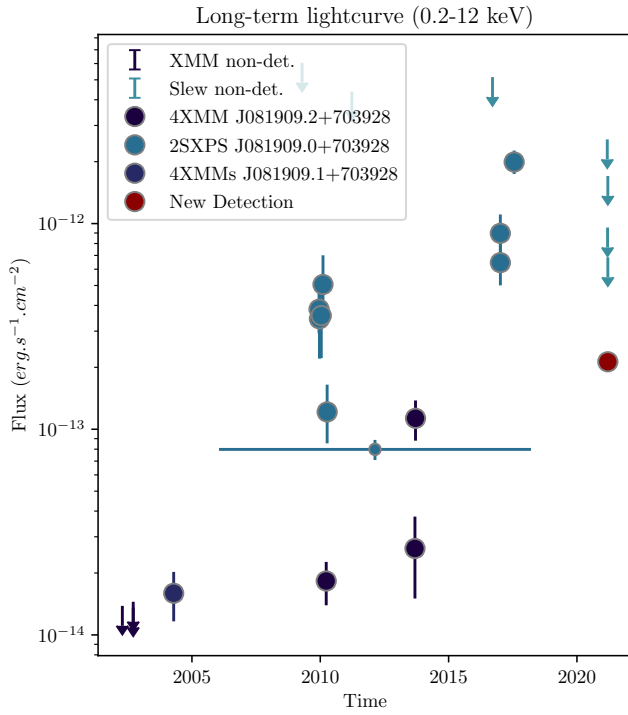


Fig. C.5. Example of an alert sent out by STONKS: a stellar flare.





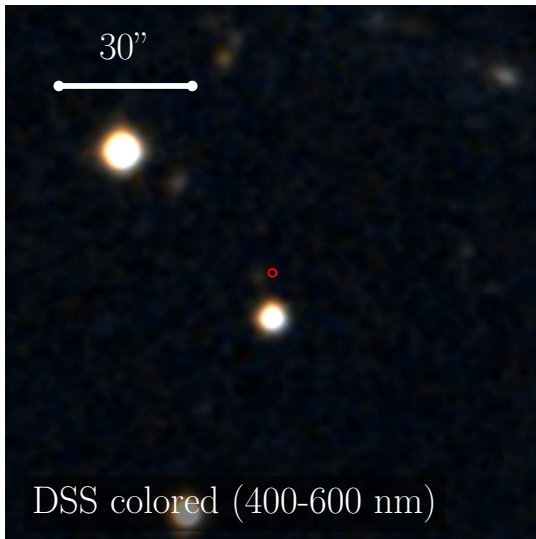
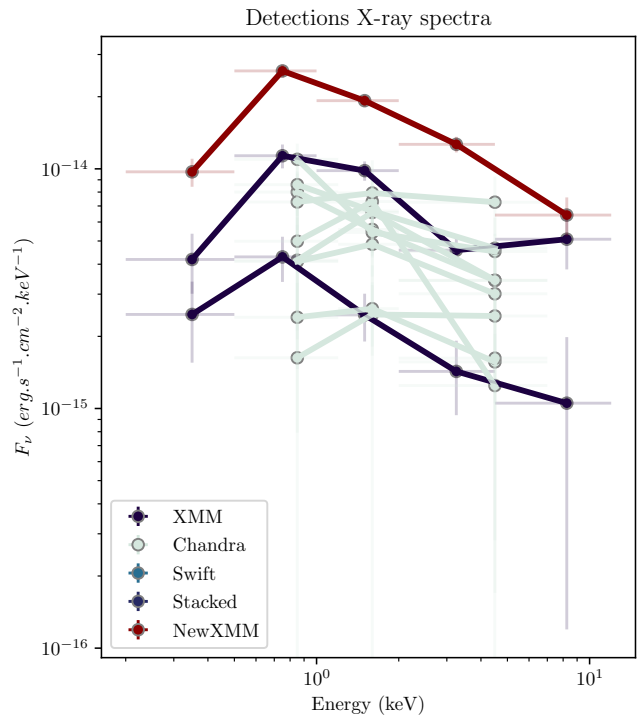
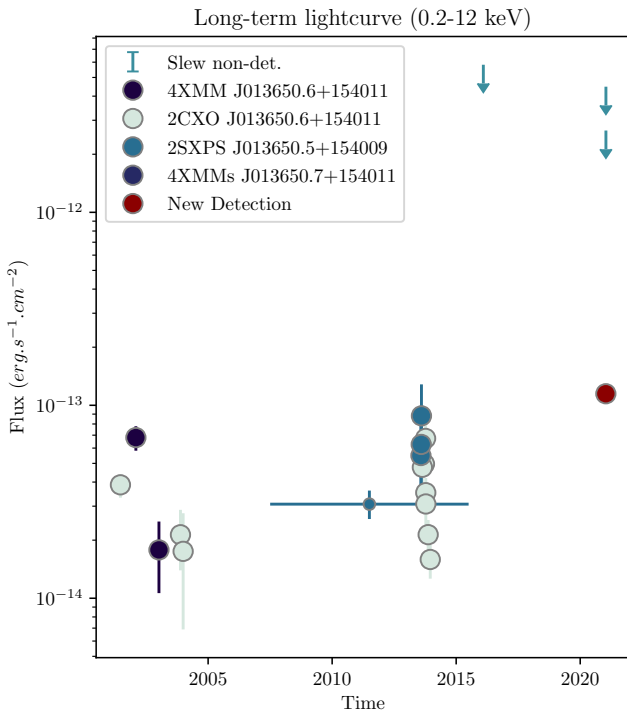
ObsID ..... 0864550401  
 Date Obs ..... 2021-03-16T01:33:03  
 Target Name ..... Ho II X-1  
 Exposure Time ..... 20357 s

SRCNUM ..... 2  
 Source RA ..... 124.79 / 08:19:09  
 Source Dec ..... 70.66 / +70:39:28  
 Position Error ..... 0.27"  
 Off-axis Angles ..... PN: 3.1', M1: 3.1', M2: 2.1'  
 Instruments DetML .....  
 PN: 1351.2, M1: 328.3, M2: 398.1, EP: 2073.8

Type of Alert ..... Past Variability  
 Long-term Variability ..... 9.9  
 Short-term Variability ..... False  
 Simbad ..... Unknown ([KWB2002] 12)

!!! Warning: extreme spectrum might impact variability !!!

Fig. C.6. Example of an alert sent out by STONKS: a possibly mis-classified ULX candidate.



```

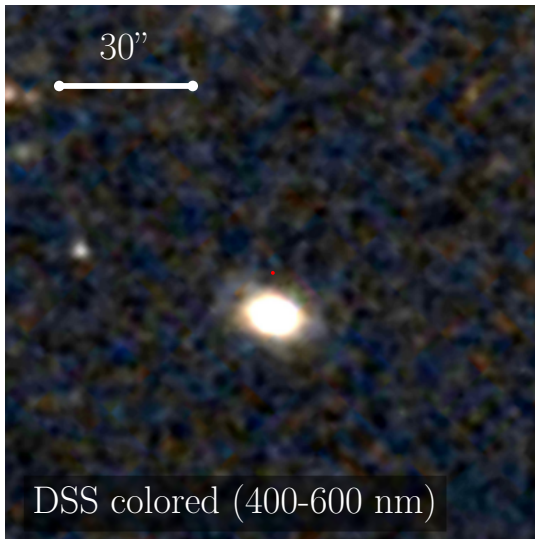
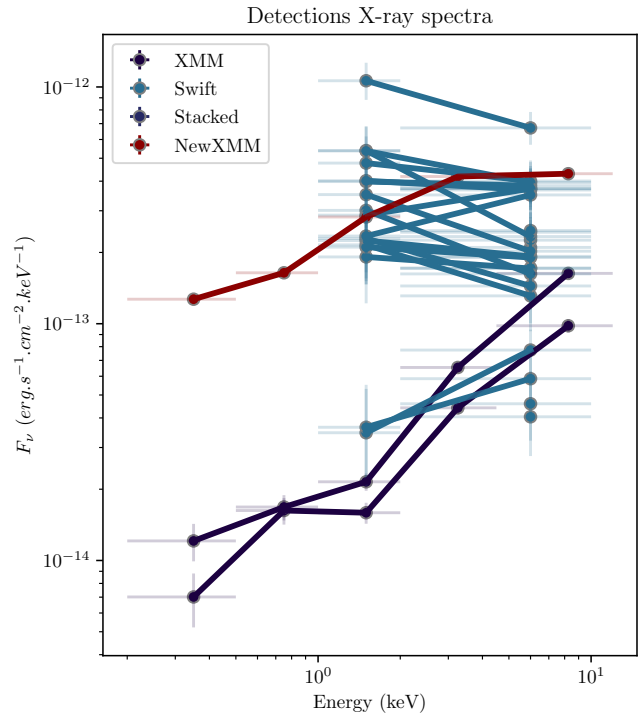
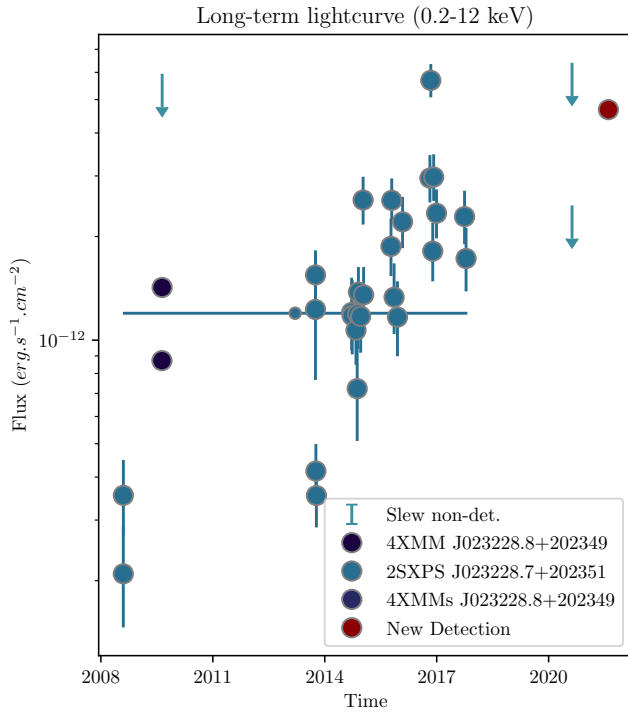
ObsID ..... 0864270101
Date Obs ..... 2021-01-13T05:44:14
Target Name ..... NGC628
Exposure Time ..... 58453 s

SRCNUM ..... 3
Source RA ..... 24.21 / 01:36:50
Source Dec ..... 15.67 / +15:40:10
Position Error ..... 0.29"
Off-axis Angles ..... PN: 6.7', M1: 7.0', M2: 6.1'
Instruments DetML .....
PN: 1334.4, M1: nan, M2: 538.9, EP: 1871.8

Type of Alert ..... High Flux State
Long-term Variability ..... 6.2
Short-term Variability ..... False
Not in Simbad
    
```

!!! Warning: extreme spectrum might impact variability !!!

Fig. C.7. Example of an alert sent out by STONKS: a new candidate XRB.



ObsID ..... 0810821801  
 Date Obs ..... 2021-08-08T23:27:55  
 Target Name ..... 1ES 0229+200  
 Exposure Time ..... 112357 s

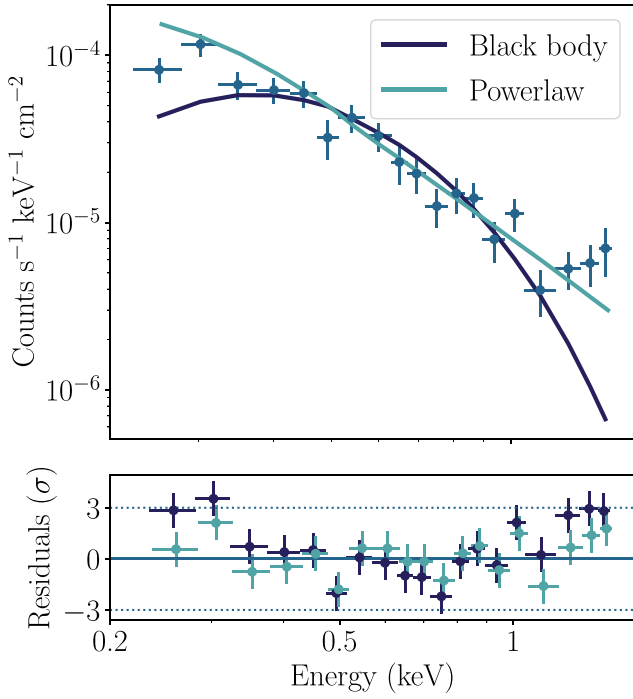
SRCNUM ..... 2  
 Source RA ..... 38.12 / 02:32:28  
 Source Dec ..... 20.4 / +20:23:49  
 Position Error ..... 0.07''  
 Off-axis Angles ..... PN: 7.9', M1: 7.9', M2: 6.8'  
 Instruments DetML .....  
 PN: nan, M1: nan, M2: 50835.3, EP: 50835.3

Type of Alert ..... Past Variability  
 Long-term Variability ..... 16.1  
 Short-term Variability ..... False  
 Simbad ..... Star (GSC2 N33123012274)

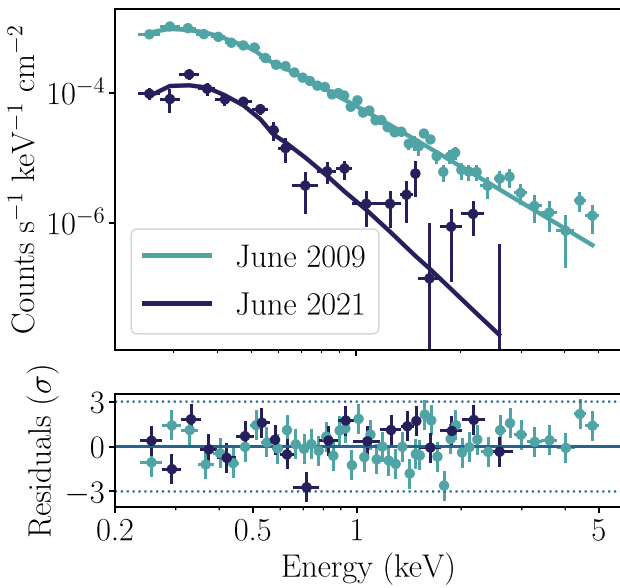
!!! Warning: extreme spectrum might impact variability !!!

Fig. C.8. Example of an alert sent out by STONKS: a short-term variable AGN with ionized absorption.

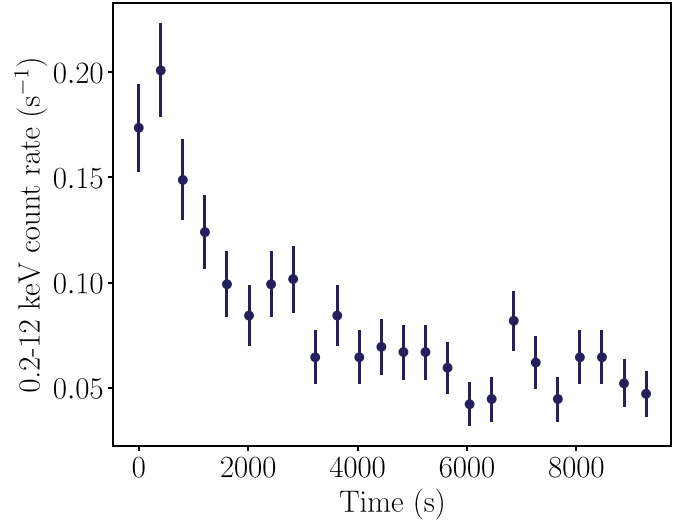
Appendix C.3. Spectra from sources of interest from the 2021 STONKS test run



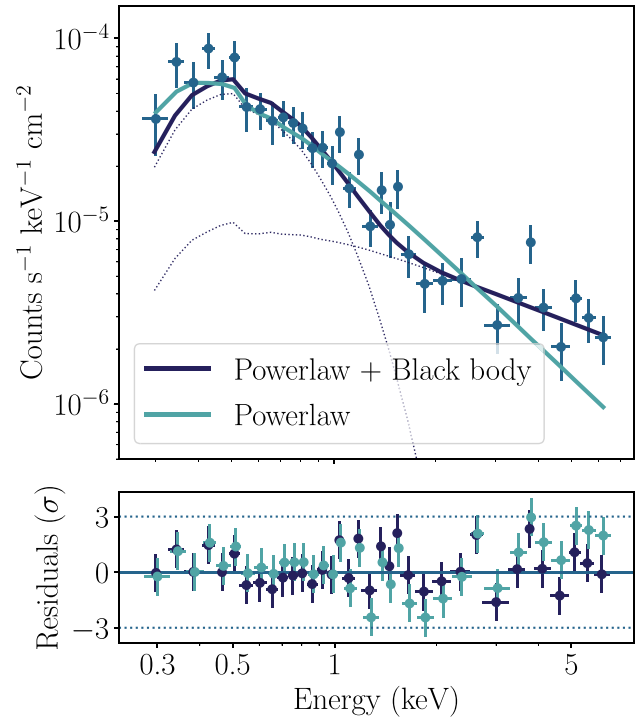
**Fig. C.9.** *XMM-Newton* EPIC pn spectrum of the TDE-like flare of 4XMM J151509.5+561347, with two models (absorbed powerlaw or absorbed black body).



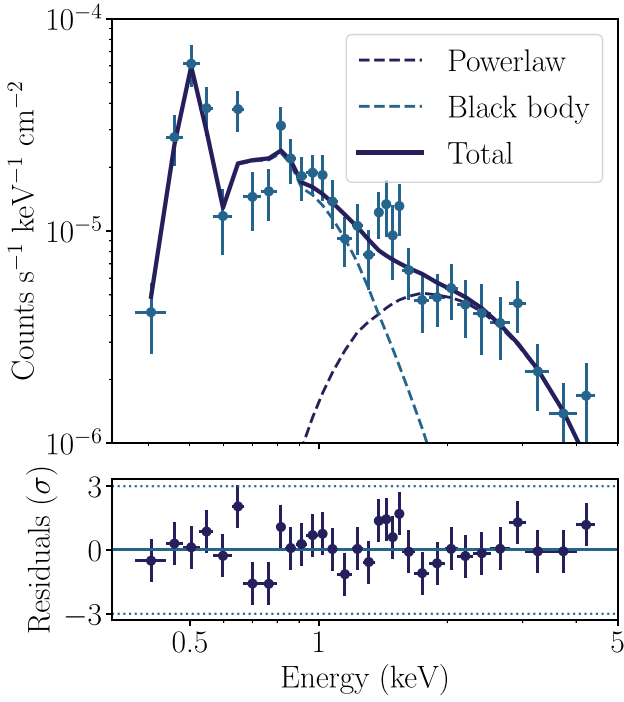
**Fig. C.10.** *XMM-Newton* EPIC pn spectrum of the variable quasar 4XMM J000532.8+200717, fitted with an absorbed powerlaw model.



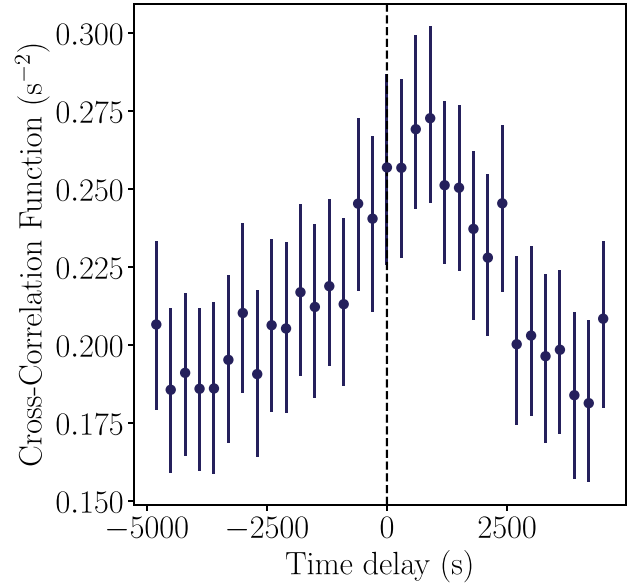
**Fig. C.11.** *XMM-Newton* EPIC pn 0.2-12 keV lightcurve of the flaring star 4XMM J053231.0+170504.



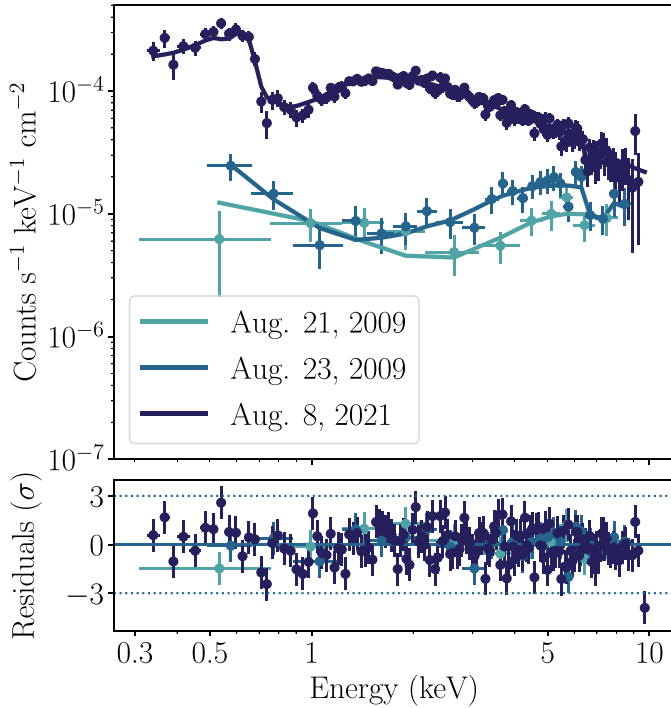
**Fig. C.12.** *XMM-Newton* EPIC pn spectrum of 4XMM J081909.2+703928



**Fig. C.13.** *XMM-Newton* EPIC pn spectrum of 4XMM J013650.6+154011

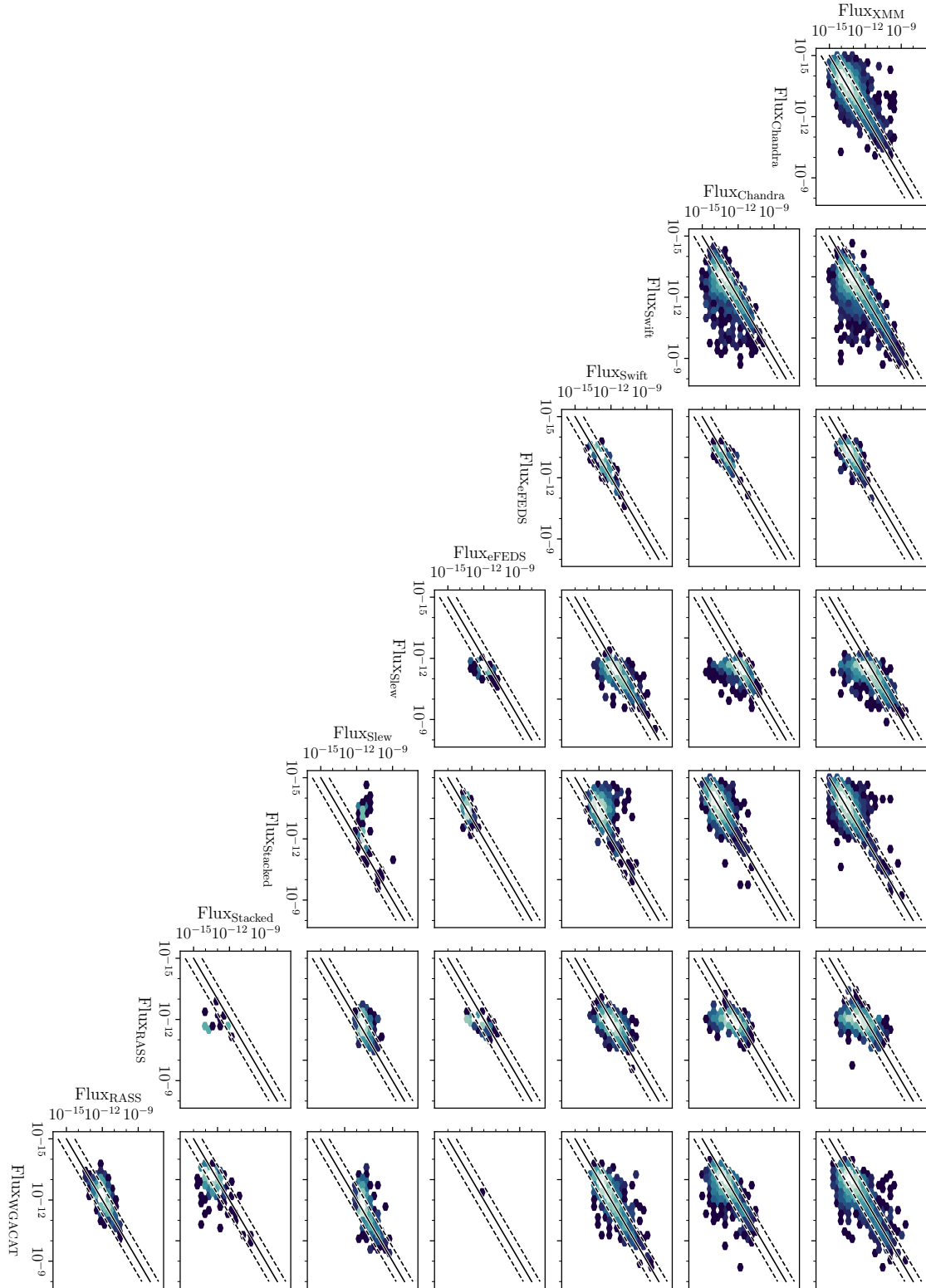


**Fig. C.15.** Cross-correlation function of 4XMM J023228.8+202349 (ObsID 0810821801), showing the lag between the soft (0.3–2 keV) and hard (2–7 keV) bands. The cross-correlation function corresponds to  $CCF(\tau) = \langle (F_{\text{soft}}(t + \tau) - \bar{F}_{\text{soft}}) \times (F_{\text{hard}}(t) - \bar{F}_{\text{hard}}) \rangle$  (e.g., [White & Peterson 1994](#)). The lightcurves were binned at 300s.



**Fig. C.14.** *XMM-Newton* EPIC pn spectra of 4XMM J023228.8+202349 from three different observations. The spectra show a variable powerlaw emission, with ionized absorption and reflection.

Appendix D: Flux comparison between matched catalogs



**Fig. D.1.** Two-by-two flux comparisons of the various catalogs within our archival cross-matched catalog. Each flux value is the average over all the catalog-specific detections, to avoid the bias towards variable sources being more observed. All fluxes are given in erg s<sup>-1</sup> cm<sup>-2</sup>, after extrapolation to the 0.1–12 keV band as explained in Sect. 2.2.