



PROJECT FINAL REPORT

Grant Agreement number: 607452

Project acronym: EXTRAS

Project title: Exploring the X-ray Transient and variable sky

Funding Scheme: Collaborative Project

Period covered: from 2014, January 1st to 2016, December 31st

Name, title and organisation of the scientific representative of the project's coordinator:

Andrea De Luca, PhD, INAF/IASF Milano

Tel: +390223699329

Fax: +39022666017

E-mail: deluca@iasf-milano.inaf.it

Project website address: www.extras-fp7.eu

Table of contents

1. Executive summary	3
2. EXTraS: context and objectives	4
2.1 The EXTraS consortium	6
3. Main achievements and results.....	8
3.1 Characterisation of short-term aperiodic variability	10
3.2 Search for pulsations.....	14
3.3 Search for new transients.....	18
3.4 Characterization of Long-term variability	21
3.5 Screening and validation of results.....	25
3.6 Multiwavelength characterisation and phenomenological classification	28
3.7 Archive management.....	30
3.8 The EXTraS Science gateway	31
4. Dissemination activities and impact of EXTraS	33
4.1 Dissemination activities.....	33
4.2 EXTraS Education Activity (EXEAt)	33
4.3 EXTraS Science Workshop	36
4.4 Publications	37
4.5 Potential impact of EXTraS.....	40
Annex	43

1. Executive summary

A wide diversity of astrophysical phenomena - from stellar flares in the solar neighborhood to accretion in galactic nuclei at cosmological distances - are characterized by flux and spectral changes on time scales ranging from a fraction of a second to several years. Every day, observing facilities with time-resolved imaging capabilities collect huge amounts of potentially interesting information, which remains mostly unused, stored in data archives. This is especially true in the high energy range of the electromagnetic spectrum, where source variability is very common but the time dimension is seldom systematically exploited.

The EXTraS project – “*Exploring the X-ray Transient and variable Sky*” – extracted all temporal domain information buried in the whole database collected by the EPIC cameras onboard the XMM-Newton mission, the most powerful tool to study variability of faint sources in the soft X-ray sky. This included a search and characterisation of all kind of variabilities, both periodic and aperiodic, in hundreds of thousands of sources spanning more than eight orders of magnitude in time scale and six orders of magnitude in flux, as well as a search for fast transients, missed by standard image analysis. Phenomenological classification of variable sources, based on X-ray and multiwavelength information, has also been performed. All results and products of EXTraS have been released to the scientific community through a web public data archive, together with new analysis tools.

EXTraS is the most comprehensive search for, and characterization of variability, on the largest ever sample of soft X-ray sources. An enormous scientific discovery space is made available to the community, with impact on the study of virtually all astrophysical source classes. Our project will certainly unveil new and unexpected classes of sources (as has always been the case when a new region in parameter space has been explored), and will raise new questions in high-energy astrophysics, enhancing the potential of discovery of the XMM-Newton mission, in itself the most productive observatory of the European Space Agency. Our variable source catalogue will trigger reanalysis of other databases as well as new observations and will become a reference in the forecoming era of large surveys, also serving as a pathfinder for future missions.

EXTraS is a collaborative effort of six European partners: Istituto Nazionale di Astrofisica (INAF, Italy, coordinator); Scuola Universitaria Superiore IUSS Pavia (Italy), Consiglio Nazionale delle Ricerche – Istituto di Matematica Applicata e Tecnologie Informatiche “E. Magenes” (CNR-IMATI, Italy); University of Leicester (UK); Max Planck Gesellschaft zur Foerderung der Wissenschaften – Max Planck Institut fuer extraterrestrische Physik (MPG-MPE, Germany); Friedrich-Alexander Universitat Erlangen-Nuremberg – Erlangen Center for Astroparticle Physics (ECAP, Germany).

2. EXTraS: context and objectives

Variability pervades the cosmos. Almost all astrophysical objects, from stars in the surroundings of the solar system, to supermassive black holes in the nuclei of very distant galaxies, display a distinctive variability – their flux and spectral shape changing on a range of time scales. This is especially true in the high-energy range of the electromagnetic spectrum. The X-ray and gamma ray sky is extremely dynamic and new classes of objects, some of them completely unexpected, have been discovered in the last decades thanks to their peculiar variability. Examples of transient, or highly variable high-energy sources are:

- gamma-ray bursts (GRBs): the most powerful cosmic explosions, likely produced by the collapse of massive stars to black holes or by the coalescence of two neutron stars;
- soft gamma-ray repeaters (SGRs): X-ray sources believed to be powered by magnetars, i.e. neutron stars with the strongest magnetic field in the Universe;
- (transient) X-ray binaries: black holes, neutron stars or white dwarfs accreting matter from their stellar companion;
- stellar flares: X-ray flares from magnetically active, late-type stars, either isolated or in binary systems;
- blazar flares: gamma-ray flares produced by the jets of supermassive black holes at the centre of galaxies
- tidal disruption events: gravitational capture and disruption of a star by a supermassive black hole;
- supernova X-ray flashes: produced by the supernova shock emerging from the exploding star.

Crucial information is often carried by periodic variability, arising from the rotation of a (compact) star, or the orbital motion in a binary system. Examples of high-energy pulsators are:

- spinning up and down accreting magnetic neutron stars in binary systems;
- spinning down young neutron stars, the emission of which is powered by the dissipation of rotational, thermal or even magnetic energy (as in the cases of classical radio pulsars, the so-called “Magnificent Seven” neutron stars, and magnetars, respectively);
- accreting magnetic white dwarf systems, such as polars and intermediate polars;
- orbital modulations (including periodic dips and eclipses) of the X-ray flux in various classes of X-ray binaries with accreting neutron stars, black holes or white dwarfs (especially if seen from a high inclination).

Variability is key to understanding the sources’ nature and physics. It is plainly impossible to summarize in a few lines the range of science topics that can be accessed and addressed by time domain investigations in the X-ray range. X-ray variability yields unique insights in to accretion physics (e.g. radiation efficiency of accretion flows; mechanisms for generating winds and jets) and strong gravity physics (e.g. conditions in the inner disk), thanks to observations of AGN, tidal disruption events and gamma-ray bursts (marking the birth of a black hole!). We can learn about the mechanisms behind massive star explosions, as well as about the progenitors of supernovae, by observing SN shock breakout events (which would also enable more sensitive searches for the long-sought-after associated gravitational waves and neutrinos). X-ray variability allows us to focus on the physics of magnetic field generation and dynamics in compact objects (e.g. through observations of violent and less violent events related to extreme magnetic fields of magnetars) as well as in normal stars (observation of stellar flares and coronal emission). The latter point holds great promise for our understanding of planetary system formation and evolution (the effects of flares on protoplanetary disks and on the habitability of planetary systems), as well as for understanding our own Sun.

Observing variability at high energy. Most of the variable phenomena mentioned in the previous section have been discovered with large field-of-view (FoV) instruments operating at hard X-ray/gamma-ray energies, which, constantly observing large portions of the sky, can also detect relatively rare events. In the soft X-ray energy range (0.2-12 keV), focusing telescopes are much more sensitive than wide-field instruments. The current generation of space observatories collect each day a very large amount of data about serendipitous sources located within their field of view, including a huge amount of information regarding their variability. Data archives from such telescopes have great potential for studying variability of (serendipitous) X-ray sources, only limited, in principle, by photon statistics and by the intrinsic time resolution of the instruments. However, such information remains mostly unused.

In particular, the European Photon Imaging Camera (EPIC) instrument onboard the European Space Agency mission XMM-Newton is the most powerful tool to study the variability of faint X-ray sources, thanks to its unprecedented combination of large effective area, good angular, spectral and temporal resolution, and large field of view. 16 years after its launch, EPIC is still fully operational and its immensely rich archive of data keeps growing. Large efforts are ongoing, to explore the serendipitous content in XMM data. Indeed, the catalog of serendipitous sources extracted from EPIC observations, dubbed 3XMM, is the largest and most sensitive compilation of X-ray sources ever produced, listing more than 500,000 detections over about 800 square degrees of the sky. Another 20,000 sources have been identified in the so-called XMM Slew Survey (XSS), using data collected while the telescope is moving from one target to the next one, having a shallower sensitivity, but covering more than 70% of the sky. Time-domain information on such a large sample of sources remains, however, largely unexplored (see Fig. 2.1).

The EXTraS project – Exploring the X-ray Transient and variable Sky - aims at fully investigating and disclosing the serendipitous content of the EPIC database in the time domain, and to make it available, and easy to use, to the whole community. Figure 2.1 shows the expected impact of EXTraS in extending the investigation of the fundamental parameter space, the variability time scale vs. count rate plane. EXTraS includes different lines of analysis, with the following main objectives:

- i. detection and quantitative characterization of aperiodic variability in the largest possible number of sources from the 3XMM catalogue, on all time scales ranging from the duration of an observation, to the instrument time resolution;
- ii. detection and characterization of the largest possible number of X-ray pulsators. For the same sample of sources as for objective i), pulsations on a period range from 0.2 s up to the largest value allowed by the duration of the observation will be searched;
- iii. detection of the largest possible sample of X-ray transients. A consistent, systematic analysis of the full dataset will allow to unveil fast, faint transients that are only above detection threshold for a very short time interval and thus missed by standard analysis and not listed in 3XMM;
- iv. detection and characterization of long-term variability, taking advantage of the large number of overlapping observations performed at different epochs. Starting from both pointed and slew data, detections and upper limits will be combined in long-term light curves spanning up to 15 years;
- v. multiwavelength characterization of all newly discovered sources (mostly fast transients), cross-correlating their positions with existing catalogues;
- vi. phenomenological classification of all variable sources. Basically, using a series of pre-defined “features” (temporal, spectral, as well as multi-wavelength properties), the probability that a source belongs to a “group” will be calculated;
- vii. compilation of a variable source catalogue. All results and products (from all points above) will be included in a public catalog that will be released to the community. A particular effort

- will be devoted to the quality control. A web site will be deployed with an easy-to-use interface, supporting as far as possible Virtual Observatory data access standards for querying online source catalogue and time-domain data;
- viii. release of public software tools. New tools, both dedicated to EPIC data analysis and of more general use, will be distributed to the community, allowing users to perform customized time-domain analysis of their own proprietary data or of archival data;
- ix. implementation of an experimental didactic program in selected high schools in the countries of partner institutions, aimed at directly involving students in our research program, also allowing assessment of the potential of a new form of citizen science.

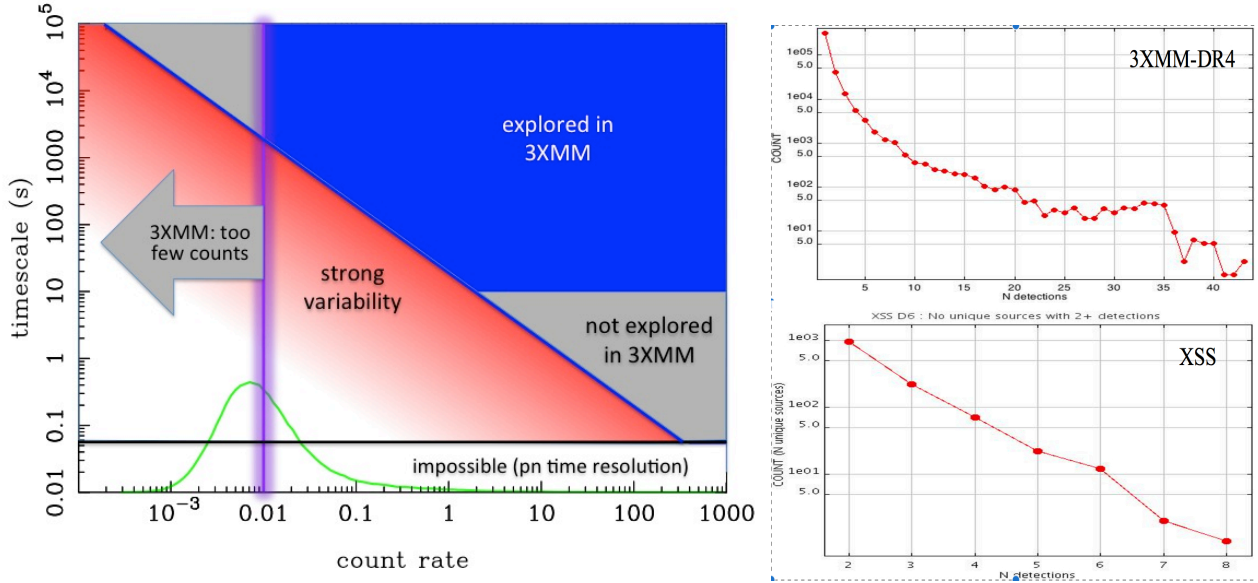


Figure 2.1. (left) The diagram shows the timescale - count rate parameter space. The diagonal blue line indicates the basic prescription (20 counts per time bin) for EPIC light curve generation in the processing pipeline for 3XMM. Above and to the right of this line the light curves exist and are characterised through simple variability parameters and flags. The vertical violet line indicates the approximate limit for light curve creation (300 counts total). To the left of this line, light curves are typically not created, except for the longest exposures. The two grey areas are not currently accessible through the 3XMM light curves, either because the smallest time binning is limited to 10 sec (for the high count rate region) or because of the 300 counts total limit (region at faint count rates). The green curve overlaid on the diagram shows the actual distribution of EPIC count rates for sources in the 3XMM catalogue. The part of parameter space not currently explored/explorable with the 3XMM light curves is shaded red. In this region one is working with time bins that are too small for a statistically significant signal for \sim steady sources, but of course for a source that has strong variability, e.g. a factor 10-100 flare for a short interval, this can be detected. The region extends further to the lower left as the variability amplitude increases. **(right)** The two plots show the large number of 3XMM sources (upper panel) and of slow survey sources (lower panel) that have been detected more than one time, due to the overlap of different observations. No systematic investigation of long-term variability has ever been performed using such information.

2.1 The EXTraS consortium

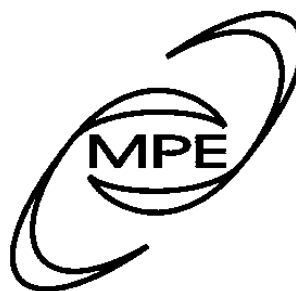
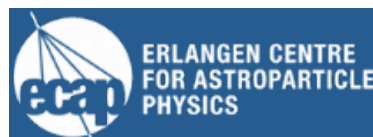
EXTraS is a collaborative effort of six European partners: Istituto Nazionale di Astrofisica (INAF, Italy, coordinator); Scuola Universitaria Superiore IUSS Pavia (Italy), Consiglio Nazionale delle Ricerche – Istituto di Matematica Applicata e Tecnologie Informatiche “E. Magenes” (CNR-IMATI, Italy); University of Leicester (UK); Max Planck Gesellschaft zur Foerderung der Wissenschaften –

Max Planck Institut fuer extraterrestrische Physik (MPG-MPE, Germany); Friedrich-Alexander Universitat Erlangen-Nuremberg – Erlangen Center for Astroparticle Physics (ECAP, Germany).

The EXTraS consortium mobilizes the best available expertise at the European level to carry out such a huge, systematic analysis of data, with subsequent classification of results, which should be made fully exploitable for a broad community. This requires a broad knowledge of variable astrophysical phenomena, a deep knowledge of the instruments, extensive experience in data analysis, in software development, in management of large databases, in high computing facilities, and in visualization tools.



**University of
Leicester**



3. Main achievements and results

A concise summary of the main achievements of the EXTraS project is given in the list below. Work performed within EXTraS in different lines of analysis is described in more detail in the following subsections 3.1-3.10.

1. We produced a thorough characterization of aperiodic, short-term variability (on time scales shorter than the exposure time) for more than 420,000 point sources included in the 3XMM catalogue. This is based on modelling of time-averaged properties of point sources in 3XMM and on careful modelling and characterization of the variable EPIC background noise. For each source we generated: (i) background-subtracted light curves with uniform time binning at 500s, ‘optimal’ and 5ks; (ii) background-subtracted light curves with adaptive time binning, based on the Bayesian block approach, with segmentation at both the 3σ and 4σ threshold; (iii) power spectra. Starting from these products, we computed a set of synthetic parameters quantifying different aspects of each source variability. We run a simplified version of the pipeline to extract light curves for the same set of sources in three energy sub-ranges and to generate hardness ratios. A set of simulations and statistical tests have been used to check and validate our products and results.
2. We systematically searched for periodic modulations on more than 300,000 sources in the 3XMM catalogue, running a pipeline based on a generalization of the FFT approach accounting for non-Poissonian noise components. For each detected signal, a refined search is performed using the folding technique; different parameters are computed (e.g. significance level, pulsed fraction) and several products are generated (e.g. light curves, folded light curves, power spectra, periodograms). If no pulsations are found, the 3.5-sigma upper limit to the pulsed fraction is evaluated. Statistical tests have been performed to check the validity of the analysis and its sensitivity. More than 60 new pulsators (and counting) have been discovered so far, with very interesting results including *the most luminous accreting pulsar ever observed*.
3. We ran a blind search for transients and highly variable, faint sources. Two approaches have been implemented. In the first one, a source detection is run on short time intervals of uniform length. In the second one, promising time intervals of optimized duration are spotted by searching for count rate changes (using a Bayesian block approach) in spatially-independent portions of the field of view, and a standard source detection is then performed on the selected intervals. Different runs have been carried out on the whole sample of EPIC observations using different pipelines with different settings. Cross-check and statistical analysis of results, together with a complete visual screening, allowed us to identify a robust sub-sample of 136 short-duration, highly-significant transient sources not listed in the 3XMM catalogue. Very peculiar events have been already spotted in such sample.
4. We systematically investigated long-term variability (LTV, on time scales longer than XMM exposures) in all detected EPIC sources, both from pointed and slew observations. The analysis was performed in three different energy ranges (total, soft, hard) and was based on (i) an improved slew data processing pipeline, resulting in an updated slew survey catalogue; (ii) a consistent computation of upper limits in slew and pointed data; (iii) collation of slew and pointed photometry, together with upper limits, and extraction of long-term light curves; (iv) search for, and characterization of variability on the resulting, typically very sparse, time series. Particular attention was devoted to the study of the compatibility of flux measurements in slew and pointed data. The main output was a LTV catalogue including more than 2 million photometric measurements for more than 420,000 unique sources, together with

meta-data for the observations used, quality information and a number of variability parameters that gauge the level and timescales of variability.

5. A screening tool was designed, implemented, tested and released, to perform an easy and complete visualization of all results and products generated by the different lines of analysis (images, light curves, variability parameters, etc). The tool also allows the user to reproduce some of the products like light curves and images, using different settings. The tool was used to perform visual screening of large samples of results, spotting, classifying and understanding different pathological effects leading to poor-quality products, which also allowed to devise strategies for statistical analysis of bulk data products.
6. We performed an automatic classification of 3XMM sources based on X-ray and multiwavelength (MWL) properties. X-ray colour information for 3XMM sources was extracted, based on the computation of hardness ratios between standard 3XMM bands. Automatic fitting of a set of models was performed on about 140,000 spectra of 3XMM sources. Automatic SED fitting was performed on all sources for which robust MWL information was available, based on a cross-match of 3XMM with 60 catalogues using the CDS XMATCH service. A training sample was defined, including about 2,900 sources in 8 different classes. Classification of 3XMM sources using the Random Forest algorithm was completed, based on X-ray spectral and MWL information, resulting in a classification error of 8.8%. An improved classification including X-ray temporal properties is at an advanced stage and is expected to be completed by 2017, June.
7. We developed an automatic pipeline to detect transient sources in Swift/XRT data in near real time and to perform an automatic classification of their nature, based on a sample of well studied events of different classes, in order to assess priorities for a fast follow-up. Transients detected by the XMM Slew processing pipeline are also included in the procedure. Optical/NIR follow-up of 21 transients was performed with the GROND instrument at the ESO/MPG 2.2m telescope.
8. The *EXTraS Public Data Archive*, including results and products for all lines of analysis (aperiodic short-term variability, periodicity search, new transient search, long-term variability) has been set up within the Leicester Database and Archive Service (LEDAS) at the University of Leicester. A new archive framework with an object-oriented code structure and a new DBMS software was implemented. User command interfaces to the EXTraS archive framework have been defined. The archive interface supports very complex queries. An EXTraS Virtual Observatory (VO) Data model was developed to allow for the publications of EXTraS data on the VO. A visualization server was implemented, to provide users with a powerful facility for interactive display of all archived data and metadata. The EXTraS Public Data Archive can be accessed from <http://archive.extras-fp7.eu>.
9. We have released the source code of the software tools developed by EXTraS to perform search for, and characterization of short-term aperiodic variability, search for periodicity, search for new transients and characterization of long-term variability.
10. We implemented the *EXTraS Science portal*, a new Science Gateway, for providing search for short-term aperiodic variability, search for pulsations and search for new transients on EPIC data. Users can select their dataset from the XMM-Newton archive and run selected EXTraS pipelines via a simplified interface, with no need for the installation of any software. All jobs are managed by the portal, based on computing resources provided by the European Grid Infrastructure. As of 2017, March, periodicity searches and transient searches have been implemented. Inclusion of the aperiodic search and characterization task is in progress and is

expected to be completed by 2017, June. The EXTraS Science portal can be accessed from <http://portal.extras-fp7.eu>.

11. We implemented an educational program for High Schools (17-18 yr-old students). We adopted an Inquiry-based learning strategy, together with a peer-to-peer education approach. This global strategy both fosters the critical thinking and engages the students, who act as though they were researchers. Based on the use of EXTraS products and software tools, students are involved in the task of classifying transient source candidates. The program was implemented in 7 workshops in selected Schools in Italy, Germany and the UK. A School kit was produced, including presentations and an IT tool (a virtual machine including EXTraS software and data) together with a user manual and a cookbook for teachers; it is and will remain available through the EXTraS web site upon request. As a future extension we plan to adopt the science gateway as the main IT tool.

3.1 Characterisation of short-term aperiodic variability

We developed a modular pipeline that has then been systematically applied to EPIC data using HPC facilities. The first part of the pipeline is devoted to model sources and to optimise event selection, implicitly defining the maximal subset of 3XMM sources that we can analyse. The second part is aimed at a detailed background characterization, taking into account the high, fast time variability of the EPIC background. For the first time, our approach allows us to study sources in high-background time intervals excluded by the 3XMM analysis ($\sim 27\%$ of the total time).

Different sets of uniform bin light curves are produced for each source and for the common background, and to each source is associated an FFT spectrum. Adaptively binned light curves are also computed, obtained extending the basic Bayesian block algorithm, in order to investigate the characteristic time-scale(s) of variability. All light curves are characterised through a large set of parameters coming from statistical tests and analyses, and by fitting them with various models. A branch of WP2 pipeline, similar to the main one, was applied in 3 energy sub-bands, producing light curves in each band and hardness ratios light curves.

The user can access our results through a catalog and a huge number of ancillary products, through the EXTraS archive. We also provide the software code we developed with some documentation on tools and products.

Selection of exposures, sources, and events

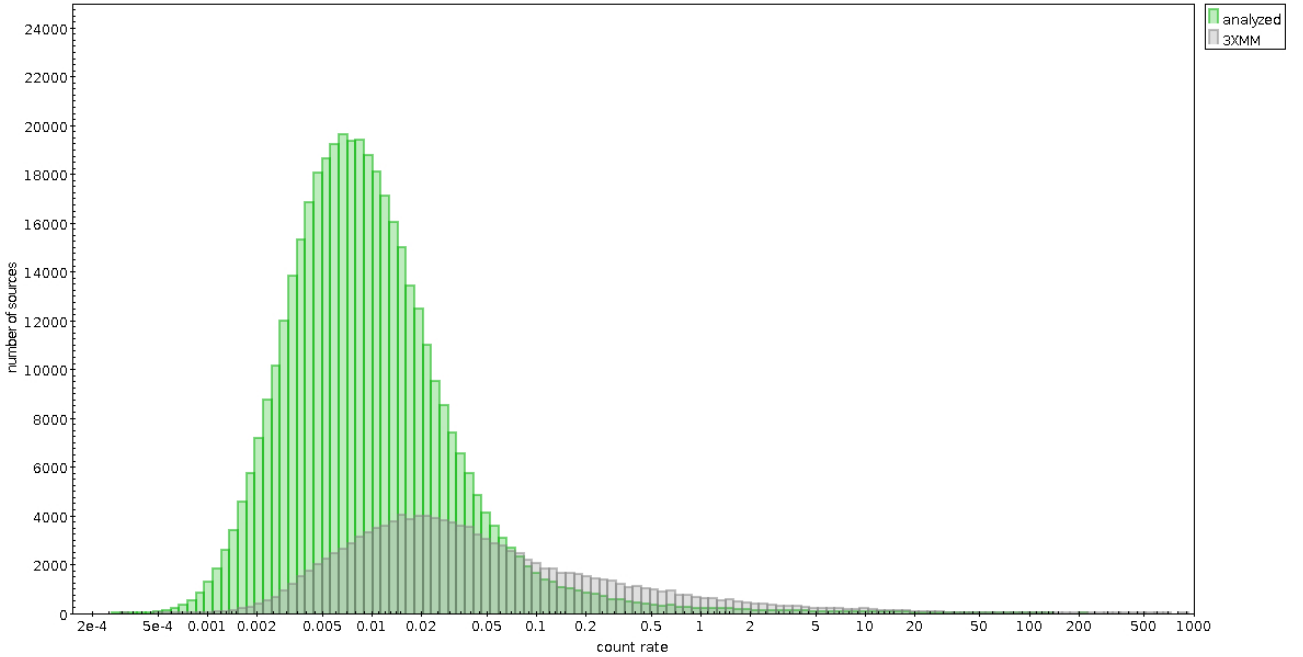
This analysis builds on top of the 3XMM-DR4 source catalog. We selected the same 21,281 exposures that were used there. Multiple (co-aligned) exposures, often collected for each camera in a single observation, were not merged - each exposure was independently studied.

We used as a starting point Pipeline Processing System (PPS) files for the MOS cameras. For the pn camera, a critical bug was discovered in PPS event files, preventing any accurate temporal analysis. Improper management of “counting mode” occurrences (due to a bug in the Science Analysis Software used for the production of PPSs) results in incoherent time tagging of events within an exposure. This forced us to produce new event files for the pn camera, starting from Observation Data Files (ODF), using standard pipeline processing with a recent release of the SAS software (SAS v14), not affected by the same bug. We applied the same quality filters to event lists as those used for the production of the 3XMM catalog; however, we considered also periods of high particle background, generally discarded in 3XMM processing. Finally we apply baricentric corrections to all the events and GTIs.

Our starting sample of sources encompasses 531,261 detections. Among the 3XMM detected sources, we select and analyse all nearly pointlike sources (angular extension $< 12''$) detected by the

3XMM pipeline, and model them in order to optimize event selection. This reduces our sample to 486,271 sources. Only sources that in a specific exposure are expected to have more than 10 collected counts are included for further analysis. We thus analyse a sample of 418,387 detected sources (802,075 if we consider multiple exposures). This number is >3 times larger than the 123,860 detections for which a light curve was generated within the 3XMM-DR4 analysis.

Figure 3.1 1: distribution of unique sources analysed within the EXTraS characterisation of aperiodic variability, compared to the one of unique sources for which a light curve was produced within 3XMM



The EPIC background

An important module of the WP2 pipeline is devoted to model and characterize the variable XMM background. Common practice has been to extract the background from a ‘background region’, independent from the ‘source region’, but with supposedly similar background properties. However, the photon background, which in our analysis includes extended sources, is far from flat. Therefore, we decided to model the background and deduce its properties in the source region. We expect different vignetting from soft proton induced background (highly variable on time scales shorter than an EPIC exposure) and cosmic X-ray background plus high-energy-particles-induced background (not variable within an exposure). We thus analyse separately the two components of background: the variable and the steady one. We exclude all sources from the counts map; we smooth it and we fill the gaps through 2D-interpolation. The remaining PSF tails are taken into account through source modelization. The final two background maps (constant component and variable components) are in rate units. These maps are used to determine selection criteria for the background regions and to evaluate the time-dependent background contribution in the source region.

As background region, we consider the entire Field of View with the exception of circular regions around point-like sources. All the sources in an exposure have the same background region. Circles radii are found through the minimization a figure of merit that balances systematics, due to photon leakage from source tails into the background, and statistic (Poisson) uncertainties in the background normalization.

Finally, for each detector region, source or background, we compile a list of parameters, needed for light curve generation. During the process of background subtraction and light curve production, we also take into account the Good Time Interval on a per CCD basis: GTIs are weighted for the

expected contribution to the source and the background.

Light curves production and characterization

We built source regions by optimizing the signal-to-noise ratio. We also took into account the contamination from other sources. This allows us to disentangle the different contributions from nearby sources (few arcsecs).

For each source, for each exposure, we produce three different light curves following different approaches:

- i. we extract “Uniform Bin” (UB) light curves in 500 s and 5000 s time bins; we also produce UB curves with ‘optimal’ bin time -- having (on average) at least 25 source counts per bin, based on the count rate reported in 3XMM. Light curves with a smaller bin size (10s) were

also produced and used for generating Fast Fourier Transform power spectra: while these spectra are released to the community, these light curves are not.

- ii. we produce “Bayesian Block” (BB) light curves – these are the optimal representation of the binned, background-subtracted light curve based on the Bayesian blocks algorithm. We group the events in a grid with a minimum number of counts from the source region. This initial grid is then processed joining those that have compatible rates. Depending on the threshold for separation we generate two sets of BB Lcs (one more sensitive and one more robust). We calibrate them through Monte-Carlo simulations to evaluate the number of expected fake blocks.

- iii. we produce light curves with uniform, 500 s time binning, based on an analysis very similar to the one implemented in the 3XMM-DR4 pipeline, following their prescription for source event selection and background subtraction. Inappropriateness of such light curves in a large fraction of cases (affected e.g. by high background, low statistics, source confusion) is known from official 3XMM caveats, but we included them in our

release as a reference, based on an independent analysis and a different background subtraction algorithm.

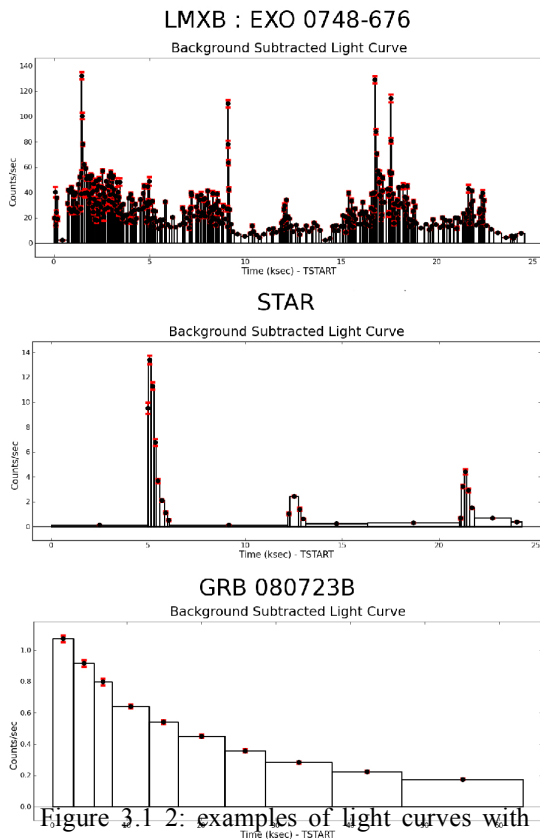


Figure 3.1: 2: examples of light curves with adaptive time binning generated using the Bayesian Block approach

We computed a large set of parameters to characterise light curves, including the weighted statistical moments of the rate distribution, the relative excess variance, the median, etc. Furthermore, a series of models are fitted to each light curve, representing global trends (like a constant, a polynomial, or an exponential decay) or local features (like flares or eclipses). Finally, we extract the cumulative distribution of the rates, and characterise it through a series of key numbers (for example the fraction of time spent at some distance from the median and the asymmetry between positive and negative deviations). Power spectra are characterised by fitting two models: a power-law plus constant and a zero-centered Lorentzian function.

Multi-band analysis and hardness ratios

An important module of the WP2 pipeline is devoted to the study of spectral variability and we run it systematically for all EXTraS sources. We considered 3 sub-bands of the full energy band (0.2-12 keV): super low (SL), 0.2- 1 keV; low (LO), 1-2 keV; high (HI), 2-12 keV. We applied similar selection criteria as for the full band: we considered point sources from which we expect at least 10 counts in an exposure. Therefore out of the 418,387 sources analysed in the full band, 221,170 were analysed in the super-low band, 197,522 in the low band, and 200,396 in the high band. We repeated in each band and for each exposure the characterisation of the background (its vignetting depends on energy). We built for each source, in each bands it has enough counts, a set of light curves, including: uniform bin light curves at 500s and optimal binning; Bayesian blocks light curves with both sensitive and robust priors. When a source could be analysed in more than one band, we built also up to hardness ratio light curves. We finally computed a basic set of parameters, such as least-square fit to a constant model and linear plus constant model.

WP2 catalog

All results have been collected in a catalog with more than 750 columns and 800,000 rows, accompanying a vast amount of products (about 17 millions files).

A total of 800,266 Uniform-Bin light curves with 500s time bin were released and 797,697 with optimal bin size. A total of 798,710 FFT spectra were released. We also produced 356,984 light curves with 500s time bin in the SL band, 338,869 in the LO band, and 322,381 in the HI band. A total of 800,993 Bayesian-Blocks light curves with a sensitive prior and 800,997 with robust prior were released.

Statistical Analysis of Results

A number of statistical tests have been performed to evaluate the consistency, robustness, sensitivity, and limits of our algorithms and results. We give here a short account only, while full details are made available through the EXTraS web site.

We investigated possible variability in light curves consistent with a constant rate and we showed that a significant flare can be missed by the fit to a constant model when the light curves have a large number of bins.

We investigated the number of negative rate bins. In light curves with 500 s time binning, it is much larger than expected from simple fluctuations: for low statistics, errors are underestimated due to the failing Gaussian assumption. As a rule of thumb, if the number of optimal bins is smaller than that of 500s bins, the user should take them and their characterisation with a grain of salt, and rely on the light curves with 5ks bins.

Thus, the user will be able to exploit different sets of parameters on the basis of the studied type of variability. Different time bins also give different information, on the basis of the time scale of the source variability: thousands 500s-time-bin light curves are compatible with a constant fit, while the corresponding optimal-time-bin are not (and vice-versa).

For Bayesian Blocks, we made a conservative assumption on the initial grid. We thus always remain in the Gaussian regime but a number of sources do not reach enough counts for a correct evaluation of variability through our Bayesian Blocks algorithm. In particular, 632,053 have a single-cell initial grid, and thus the results of the Bayesian Blocks algorithm is useless.

However, Bayesian Blocks light curves are able to detect variability at any time scale for sources with enough counts: uniformly binned light curves are often less effective in spotting localised, short features, like flares or eclipses, than Bayesian blocks. When highly-significant features appear in one kind of light curves, they are detected consistently in all light curves.

As a final example of the science that can come from our catalog, in the statistical report we used the P-value associated to a constant model for optimal bin light curves to evaluate possible difference among the Galactic and Extra-galactic population of sources, on the basis of their Galactic latitude. The two distributions appear different with the Galactic sources population more variable than the Extra-Galactic one. Although subtle biases can come into play and should be carefully evaluated, this extremely simple query on the WP2 database could potentially have important scientific implications.

3.2 Search for pulsations

The main goal of this line of analysis is the search for periodicity in the largest possible sample of 3XMM sources.

Data preparation

Our starting point are event files included in the PPS for MOS1 and MOS2 cameras. PPS event files for the pn are affected by incorrect time tagging (see Sect 3.1), possibly hampering the search for coherent signals. Thus, for the pn, we used new event files, produced within EXTraS starting from raw data (ODF) using standard tasks of the XMM-Newton Science Analysis Software (SASv15).

In most XMM observations, a single exposure is collected for each EPIC detector. However, for a number of reasons related e.g. to operational constraints, particle background, etc, more than one exposure can be collected by each detector. We merged event files resulting from such exposures for a single camera whenever quality flag were acceptable, operating mode, pointing and roll angle were consistent and the attitude was stable. Photons' time of arrival are shifted to the Solar System barycenter. For each source, we select photons from an optimized extraction region. A specific event filtering is used for sources partially laying on pn CCD edges, in order to include the maximum possible number of photons in the analysis.

Search for periodicity

To carry out the periodicity search, we adopted the algorithm described in Israel & Stella (1996). It is based on the analysis of the source power spectrum, generated using all photon times of arrival (no time binning is performed, to preserve the full information included in data). The algorithm is able to detect signals in power spectra where the statistics of the continuum might be not only dominated by white noise but also by non-Poissonian noises. These kind of noises are often observed in accreting X-ray sources and likely associated with the aperiodic variability of the source. The software includes a smoothing algorithm in order to evaluate the spectrum continuum plus a detection one which infers the main signal parameters (such as period and significance). For each candidate signal, a refined analysis is run to get a more detailed characterization: a periodogram is generated, based on the epoch folding technique; a refined value of the period is computed; a folded light curve is generated; the pulsed fraction is computed (no background subtraction is performed, thus the actual pulsed fraction of the source is higher).

If no significant peak is detected at the signal search stage, a 3-sigma upper limit to the pulsed fraction is computed.

Systematic timing analysis

The above mentioned steps are performed through an automated pipeline that we dubbed Coherence IN Epic detectors Mega Analysis and Search for COherent PERiodicities (CINEMA-SCOPE). The CINEMA pipeline is devoted to data preparation, while the SCOPE pipeline is devoted to the timing analysis.

Within SCOPE, for each single observation the search is performed multiple times for each source: this is related to the different time resolution in data collected by different cameras (e.g. pn vs. MOS), or in data collected by different CCDs in the same instrument when specific operating modes are adopted (e.g. Small Window mode in MOS). We have implemented a decision tree as follows: for each source we first perform an analysis using data with the highest available time resolution; we add subsequently data with lower resolution (if available, as in most cases) and perform a new analysis, losing in time resolution but gaining in statistics. Of course, a higher time resolution allows us to detect higher frequency signals, while increasing the statistics allows us to detect weaker signals. Thus, we carry out a number of timing analysis that optimize the signal search in different frequency intervals based on the specific time resolution in data for each single source. Temporal analysis is performed on all sources having at least 50 counts in a dataset (including combinations of data from different cameras/operating modes).

The pipeline was run both on the full dataset, as well as after rejecting time intervals affected by high background (soft proton flares). We include in our final release results and products from the run on the full dataset only.

Products and results

We give below a very concise summary of the results of runs carried out during the project:

- ~ 10 Million of Fourier analyses performed;
- ~ 3 Million of Discrete Periodic Searches;
- results produced on ~ 300,000 unique sources;
- ~ 0.15 Million candidate periodic signals detected.

The files produced as final results by CINEMA-SCOPE are:

- A light curve for each source. In the version for public release, a 400 s time binning has been used.
- A plot of the Power Spectral Density for each single 3XMM source (see Fig. 3.2.1).
- A plot of the Signal detection threshold as a function of the signal frequency for each 3XMM source above a predefined minimum number of counts (see Fig. 3.2.2)
- A plot of the Pulsed fraction upper limit as a function of the signal frequency. for each 3XMM source above a predefined minimum number of counts
- a phase-folded light curve (generated using the `efold. tool`¹) for each candidate signal found
- A plot of a periodogram for each candidate signal. It is generated by folding the data over a range of periods and by searching for the maximum chi-square against a constant in the folded profile, as a function of the period, using the `efsearch tool`²

¹ See <http://heasarc.gsfc.nasa.gov/xanadu/xronos/help/efold.html>

² See <http://starchild.gsfc.nasa.gov/xanadu/xronos/help/efsearch.html>

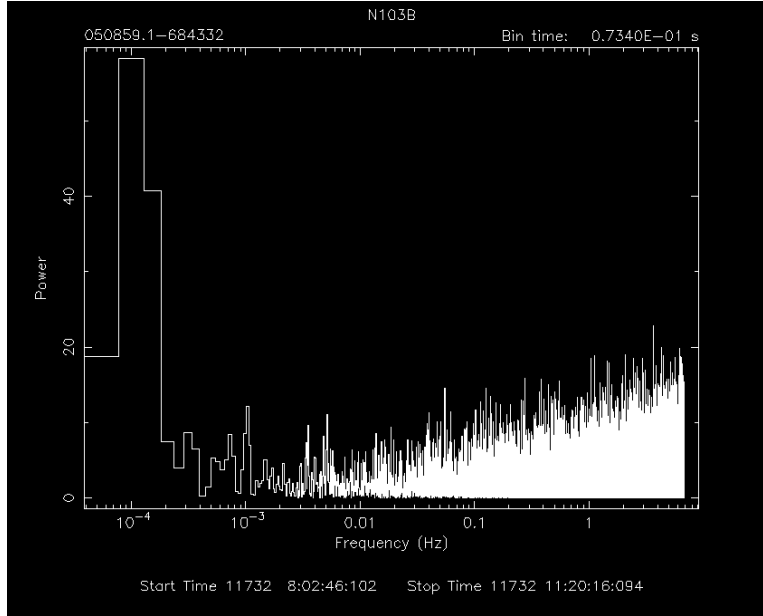


Figure 3.2.1: Plot of the Power Spectral Density for a 3XMM source

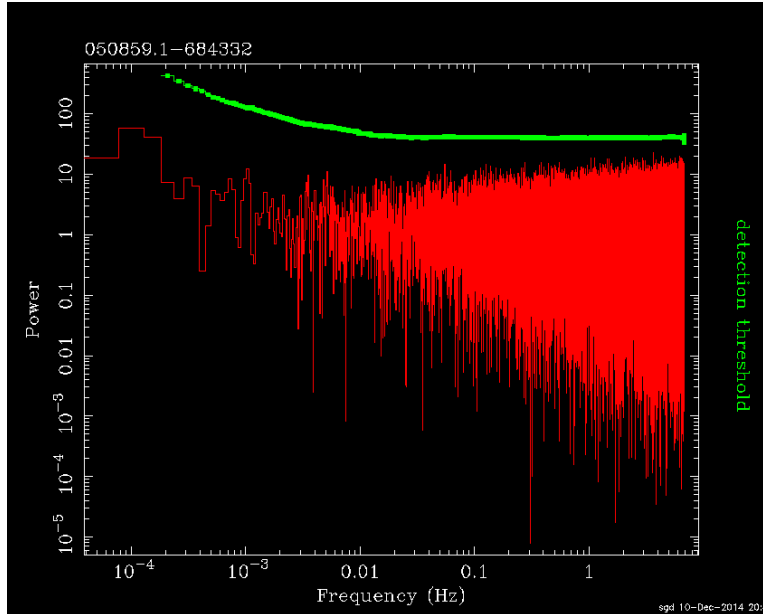


Figure 3.2.2: Plot of the Signal detection threshold (3.5σ) as a function of frequency for a 3XMM source. No significant pulsations are detected in this case.

The Catalogue

All relevant parameters characterizing each source are included in our final catalogue. We have divided the information in the catalogue into four categories:

1. Observation (OBSID) information.
2. Single source (SRC) information.
3. Information about the parameters of the periodic signals search.
4. Peaks (signals) information (for statistically significant peaks in the power spectra).

A detailed description of each column is included in the documentation pages supporting the online EXTraS public data archive.

Spurious signals in the catalogue

The distribution of all found periodic signals over the whole 3XMM data set by the CINEMA-SCOPE pipeline system is shown in figure 1. The main feature is a relatively large and high (in terms of number) peak around ~ 100 s. This is mainly due to a spurious signals related to occurrences of the ‘counting mode’ in the pn camera, a special (non-science) instrument mode where no transmission of information for individual X-ray events occurs³, triggered by strong count rates (higher than ~ 300 counts/s for the imaging modes which we use). Time gaps due to the counting mode can be up to half of the overall observing time in peculiar observations. The duration of such counting mode episodes is very variable, usually of the order of 1-2 minutes.

We checked a large sample of detections around the 100s peaks of Figure 3 and we always found that the signals are due to the (almost periodic) gaps introduced in the GTIs after correcting the time series for the counting mode issue. The spurious signals are present both in the pn and pn+MOS(s) time series FFTs.

The second relevant peak in the distribution is for large periods, above about 5,000 seconds. The corresponding signals are partly due to spurious detection due to the intrinsic aperiodic variability of the sources (affecting the low frequency part of the FFTs). We checked a large sample of these signals: though the spurious peaks are in large number there are also genuine signals. The third distribution peak in the 5-15 seconds range is dominated by XMM observations of known pulsators (mainly magnetars).

It is therefore evident that any peak reported in the WP3 catalogs with a period in the about 20-200 seconds range has to be treated with caution and a visual inspection is recommended.

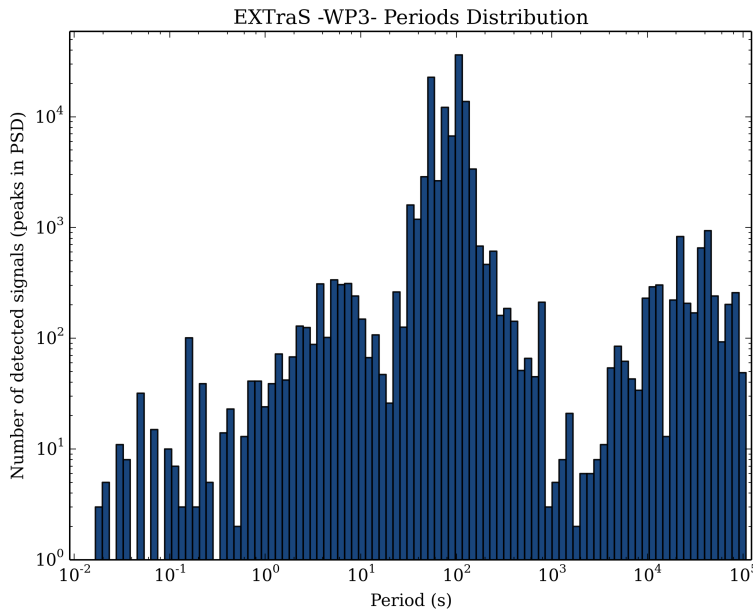


Figure 3.2.3 Distribution of found periodic signals over the whole 3XMM data set by the CINEMA-SCOPE pipeline system.

3 From <https://heasarc.gsfc.nasa.gov/docs/xmm/uhb/brightavoid.html#uhb:par:bright>

CINEMA-SCOPE scientific results

In-depth validation (requiring visual inspection) and analysis of selected signals is in progress. We have, as of 2017, February, identified more than 60 new pulsators. We have already reported in the literature on several important scientific discoveries: the most interesting one is the detection of a pulsar in the extreme Ultraluminous X-ray Source NGC5907 ULX-1, located at a distance of ~ 17.1 Mpc. Shining at a rate exceeding by a factor 1000 the Eddington limit for a neutron star, and displaying the largest spin-up rate ever observed, such source challenges current accretion models (Israel et al., 2017, *Science* 355, 817). A complete list of publication is included in Sect. 4.1.

3.3 Search for new transients

The aims of this line of analysis were the implementation and systematic application to EPIC data of an algorithm for the identification of X-ray sources which can be significantly detected in a short time interval but not by the analysis of the full observation. This happens to strongly variable sources that are too dim to emerge from the background of a long observation or that are bright enough only during periods of high particle background removed by the standard analysis.

Uniform time slicing

As a first step, we implemented a software pipeline to perform the source detection on images of fixed time duration and compare its output with the source list of the full observation, that is included in the PPS products. We define “transient candidates” all the sources detected in at least one time interval, but with no counterpart in the PPS source list.

The pipeline is driven by a master C-shell script, which runs a series of C-shell, C++ and Python scripts, most of which use HEASOFT FTOOLS and SAS tasks. In particular, the source detection in the time intervals is performed with the SAS tool `emldetect`, using the same energy bands (0.2-0.5 keV, 0.5-1 keV, 1-2 keV, 2-4.5 keV, 4.5-12 keV, and the cumulative 0.2-12 keV band), selection criteria and options as in the PPS and 3XMM catalogue.

After extensive testing (which allowed us to discover a bug in the PPS that required us to reprocess the PN event files available in XMM-Newton Science Archive), this pipeline was systematically run on all EPIC observations included in the 3XMM-DR5 catalogue (Rosen et al. 2016, *A&A* 590, A1) with time intervals of 1000 and 5000 s. This analysis required more than 45,000 computing hours on a computer cluster at CINECA, the largest Italian computing centre, and produced a very large number of transient candidates: 104,583 and 95,410 sources for the 1 ks and 5 ks time bins, respectively. Considering that only point-like sources are expected to be variable, we could select as transient candidates only sources with `EXT=0`, but their number was still very large (80,211 for 1 ks and 60,883 for 5 ks) and the manual screening of a random sample unveiled a very high fraction of spurious sources.

Adaptive time slicing

To reduce the number of spurious detections and to sample a broad range of time intervals, we designed a more effective detection technique, based on a modified version of the Bayesian blocks (BB) algorithm (Scargle et al. 2013, *ApJ* 764, 167). The BB adaptive-binning algorithm finds statistically-significant count rate change-points by maximizing the fitness function for a piecewise-constant representation of the data, starting from an event list. Our modified version can account for highly-variable background such as that found during proton flares in XMM-Newton data. For each observation, we divide the field of view in partially-overlapping $30'' \times 30''$ regions and we run the BB algorithm on each of them. Regions with no significant variability with respect to the local

background light curve return only one block covering the whole observation, while regions containing candidate transients return more blocks.

New transients nearby 3XMM sources

To properly evaluate the background light curve and to minimize the contribution from the possible variability of known sources, the BB algorithm excludes sufficiently large (depending on the source intensity) regions around the point sources detected in the full observation. To examine also these regions, where interesting transients might be hidden (especially in crowded X-ray fields, such as star-forming regions and nearby galaxies), we developed a dedicated algorithm. For each observation, it creates images integrated over a fixed time interval (e.g., 1000 s) of regions with a side of 40 arcsec around the sources excluded by the BB algorithm and tests for the presence of excesses in addition to known sources on a grid of fixed positions using a sliding cell.

Among the time intervals identified either in this way or by the BB analysis, we selected only those with duration shorter than 5 ks (the minimum duration of standard EPIC exposures) and triggered by regions with a spatial distribution of the events which are better fit (at a >5 sigma confidence level) with the addition of a point source model rather than by a simple isotropic background. We then ran the same pipeline described above using these time intervals instead of the ones with fixed duration and retained as good transient candidates only the new sources detected within the box that triggered the corresponding time interval.

The additional code required by this improved version of the pipeline is mainly composed by Python scripts and routines and was tested and optimized using computer clusters at IASF-Milano, Osservatorio Astronomico di Trieste (INAF), IUSS Pavia and Leicester University.

Systematic data analysis and results

Different versions of the pipeline were run several times on the full set of EPIC data analyzed to produce the 3XMM-DR5 catalogue and the corresponding results were screened to extend this catalogue with the addition of high confidence transient sources. From a list of several thousands transient candidates, we excluded all those with relatively low detection likelihood ($\text{DET_ML} < 15$ in the 0.2-12 keV band in all the active EPIC cameras) and the transient signals identified as bright/flickering pixels by a dedicated automatic tool. The time-resolved images and light curves of the remaining transient candidates were visually screened to exclude spurious detections (produced, for example, by out of time events or stray-light rings of bright sources, or by a bad satellite attitude reconstruction) and sources with no clear transient behavior, obtaining a final list of 136 new transients. Most of them (122) were discovered thanks to the BB algorithm and 14 through the analysis of the regions close to 3XMM sources, using 1 ks time bins.

By cross-matching the transient positions with stellar catalogues and inspecting the corresponding optical images, it becomes clear that most transients are very likely stellar flares (including flares from young stellar objects, as shown in Pizzocaro et al 2016, A&A 587, A36). This interpretation is also confirmed by the distribution of their duration (defined as the length of the time interval where the source was detected), shown in Figure 3.1.1: only a few transients are shorter than 700 s (and none shorter than 5 minutes), as expected for a population dominated by X-ray flares of active stars (see, e.g., Figure 20 in Pye et al. 2015, A&A 581, A28). The predominance of transients of Galactic origin is also suggested by the distribution of the Galactic latitude of the 136 transients (Figure 3.1.2), where the fraction of low latitude objects is much larger than in the 3XMM-DR5 point sources with $\text{DET_ML} > 15$. However, we note that the position of the shortest transient is consistent with a galaxy at redshift $z=0.093$. Due to its similarity with the X-ray transient detected by

Swift/XRT in coincidence with SN 2008D (Soderberg et al. 2008, Nature 453, 469), we interpret it as the ~ 5 minutes X-ray flare produced by the shock break-out of a supernova.

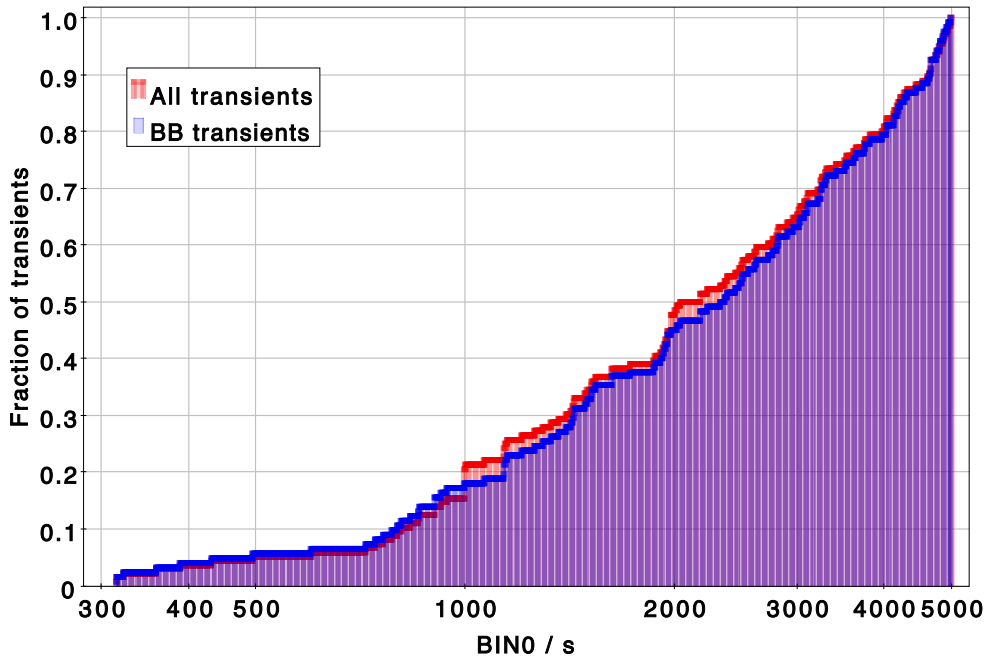


Figure 3.1.1 - Cumulative distribution of transient durations (in seconds), defined as the length of the time interval where the source was most significantly detected. The subset of transients discovered using the BB algorithm is indicated (blue) to exclude the sources discovered with the close-to-source algorithm, whose durations can only be integer multiples of 1000 s.

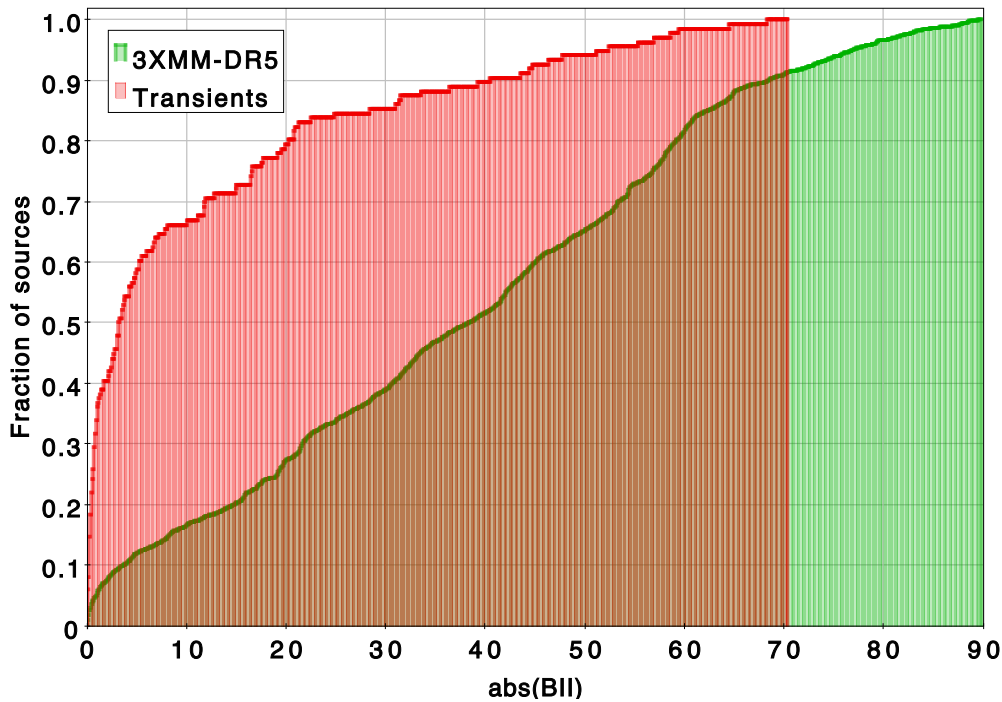


Figure 3.1.2 – Cumulative distribution of transient Galactic latitudes (absolute value), compared with the one for point sources (EP EXTENT=0) with DET ML>15 and SUM FLAG=0 in the 3XMM-DR5 catalogue. The excess of transients at low latitudes is clearly visible.

3.4 Characterization of Long-term variability

The goals of the long-term variability (LTV) work-package were to (1) collate long-term X-ray photometric data from the XMM-Newton (pointed and slew) serendipitous surveys, into a catalogue from which long-term (day to year timescale) light curves (LCs) can be constructed, and (2) test for, and characterise, variability in those long-term LCs.

Data inputs

The three main data inputs to the LTV catalogue are (i) a set of detections from pointed XMM-Newton observations, (ii) a list of detections from XMM-Newton slew observations and (iii) a set of upper limit measurements of all unique sources, obtained from all pointed and slew observations where a source position was imaged but the source was not significantly detected. XMM-Newton data used for the LTV analysis span approximately 15 years, since February 2000. While most sources have between 1 and 5 observation snapshots, dedicated observations of some regions of the sky and slew scans through the ecliptic poles result in some sources being observed up to 51 times.

The intent at the outset was to combine the existing publicly available 3XMM-DR5 catalogue of detections from pointed XMM-Newton observations with a new list of XMM-Newton slew detections. However, a simple combination was not straightforward because while photometry for the former is generated in 5 ‘narrow’ energy bands, plus a broad total, energy band, the slew data are generated in 2 broad bands, plus the total band. Due to the subtle but important differences in the way pointed and slew data are processed, and since photon numbers are usually small in slew data, to maximise consistency of the slew and pointed photometry, we recomputed pointed photometry directly in the same broad bands (0.2-2keV, 2-12 keV and 0.2-12 keV) used in slew pipeline processing.

A substantial element of the work under the EXTraS remit involved an overhaul and upgrading of the source detection components of the slew data processing pipeline software, and the consequent creation of a new list of slew sources. The improvements to the slew processing code include (i) investigation, construction and implementation of a new, FOV-average PSF model that reflects the variation of the source profile as a source passes across the field of view (FOV) during a slew observation, (ii) inclusion of an overlap region in adjacent slew sub-images to allow better treatment of sources that appear near the edges of the images, (iii) improved processing and image sub-division of long slews to minimise problems in slews that loop back in RA/DEC and (iv) more sophisticated handling of slews affected by high background and corrupted attitude information, allowing more data to be accepted.

The new slew pipeline was run on all XMM-Newton slew data available at Dec 31 2014, also taking advantage of the most recent updates in instrument calibration, yielding a new list of slew detections containing 29945 detections.

The other key component in the LTV catalogue is upper limit information. Unique sources (and their constituent detections) from the slew data were first matched, where possible, to existing unique sources from the pointed data, with residual, unmatched slew detections assigned to new unique sources. The resulting, intermediate, complete list of unique sources was then used to generate upper limits from any slew or pointed observation where the source was in the FOV but not detected above a threshold detection likelihood. The upper limits were computed via a version of the publicly available FLIX upper limit tool, adapted to the specific requirements of the EXTraS LTV analysis, i.e. computed in the same broad bands. The upper limits were added to the complete set of pointed and slew detections. We note that, as well as including these upper limits, fluxes of existing pointed

detections with detection likelihoods < 8 were replaced by upper limit data, for consistency with the threshold used for slew data (equivalent to a Gaussian significance ~ 3.4).

The complete LTV catalogue comprises 408 columns and contains ~ 2 million rows ($\sim 566,000$ pointed detections, $\sim 30,000$ slew detections and 1.43 million upper limit measurements), from 419,240 unique sources.

Variability analysis

The detection and upper limit data associated with each source form the long-term light curve of the source. A significant issue for the LTV analysis is that the XMM-Newton long-term light curves typically have few (< 5) data points and these are generally both sparsely and non-uniformly distributed in time. As such, simple approaches have been adopted to search for and characterise, variability in them. Software was developed to perform several analyses of these light curves. To provide a means of establishing potential variability in the light curves, 3 key parameters were computed: (1) the maximum-to-minimum flux ratio (and its error), including upper limit data if the minimum point was an upper limit, (2) the largest error-normalised flux change between any two detections in the light curve (which represents a measure of the significance of the change), and (3) a reduced chi-square measured about the weighted mean flux and associated probability of the null (constant source) hypothesis, based on detections. Additionally, a runs (Wald-Wolfowitz) test was performed on detection data – this measure identifies systematic runs of consecutively positive and negative deviations from the mean, which can be useful for sources with sufficient numbers of points (> 10 points).

The LTV analysis also quantifies the largest upward and downward flux change ratios, the timescales over which they occur and the shortest timescales observed in the data for factor 2 and factor 10 upward and downward transitions (where present) – these are all determined conservatively, with the change being measured between the lower and upper 1-sigma error bar limits of the relevant pair of brighter and fainter flux points.

Quality information is important so that users are aware of source light curves whose accuracy may be compromised by data issues. The LTV catalogue provides various detection-level and unique-source-level flags that users can employ to assess the quality of the light curves. In particular, for each unique source, flags are included to indicate (a) the quality flag of its lowest quality detection, (b) the fraction of its detections that are detected as extended, (c) the fraction of its detections affected by pile-up and (d) whether any of its constituent detections are affected by potential astrometry issues. It should be stressed that many sources with quality flags indicating possible issues are valid sources and their light curves may well be reliable, though the veracity of the data should always be confirmed.

Along with the LTV catalogue, which is being provided both through the EXTraS project database and as a stand-alone fits file, a set of associated graphic products, as illustrated for example in figure 1, are provided, that display the light curve for each source and the primary, above-mentioned variability measures. The graphic is available for each of the three energy bands and for each instrument and is accessible via the database and a URL within the fits file.

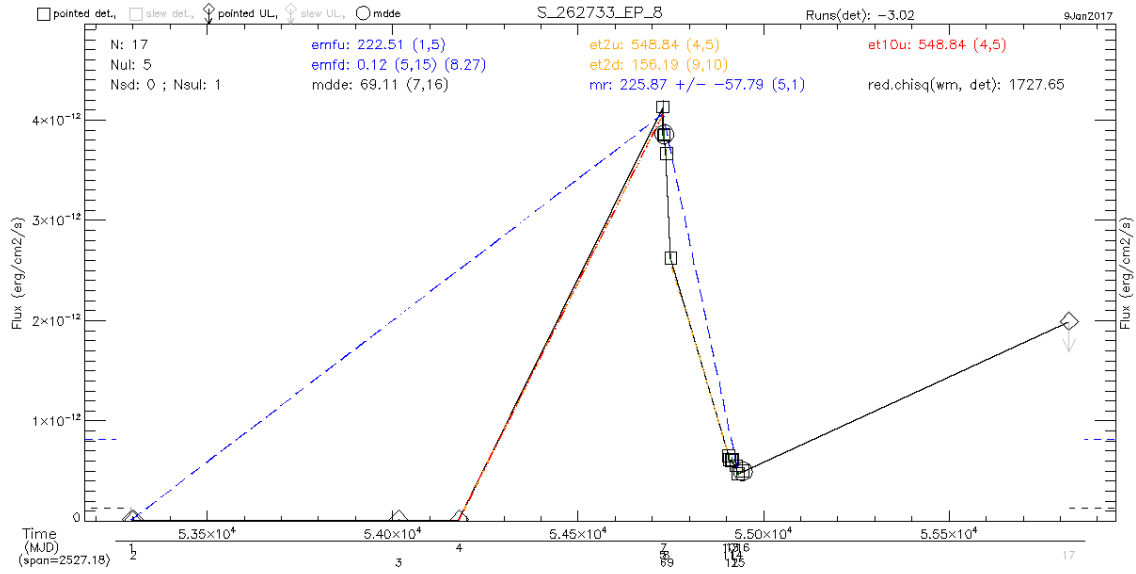


Figure 3.4.1 - Example LTV light curve graphic. Detections and upper limits are indicated by distinct symbols (for both slew and pointed data). Parameter values from the LTV analysis are included in the panel, along with lines that indicate the transitions involved.

Scientific content

The LTV catalogue has stand-alone scientific merit but can also be used in conjunction with data from the short-term aperiodic and periodic analyses.

From the $\sim 420,000$ unique sources, some 74,000 have >1 photometric measurement. Of these, $\sim 11,000$ show a flux change with significance >5 . The subset of these with the cleanest quality flags contains ~ 4900 sources. Figure 2 plots the EPIC band 8 F_{\max}/F_{\min} ratio against flux, and highlights the selection of a sample of ‘clean’, significantly variable sources.

To gauge the reliability of the data and analysis, simulations have been run. These involved generating simulated constant sources and (i) examining the properties of the key LTV parameters recovered, e.g. as a function of numbers of points in the light curve, (ii) comparing the distributions of recovered LTV parameters against those of the real sources and (iii) estimating false positive rates for the detection of variability. Results are made available through the EXTraS web site. Several small samples of sources were also manually screened to check that quality filters remove the most problematic cases and to try and identify residual issues not covered by the quality flags. The screening analysis and issues arising, e.g. occasional unresolved sources and imperfect matching of detections into sources, are described on the EXTraS web site.

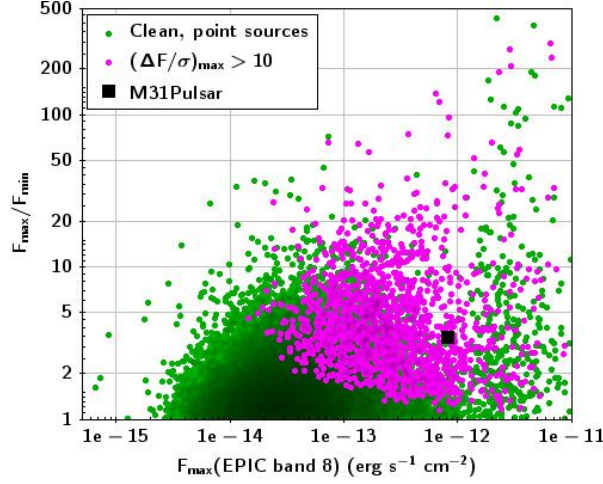


Figure 3.4.2: Distribution of sources in the F_{\max}/F_{\min} v F_{\max} plane (for EPIC, band 8). Green points are point-like sources with ‘clean’ quality flags while magenta points are the subset where the largest error-normalised flux of any pair of detections in the light curve is > 10 . The 1.2s pulsar discovered through the work of WP3 is shown in black.

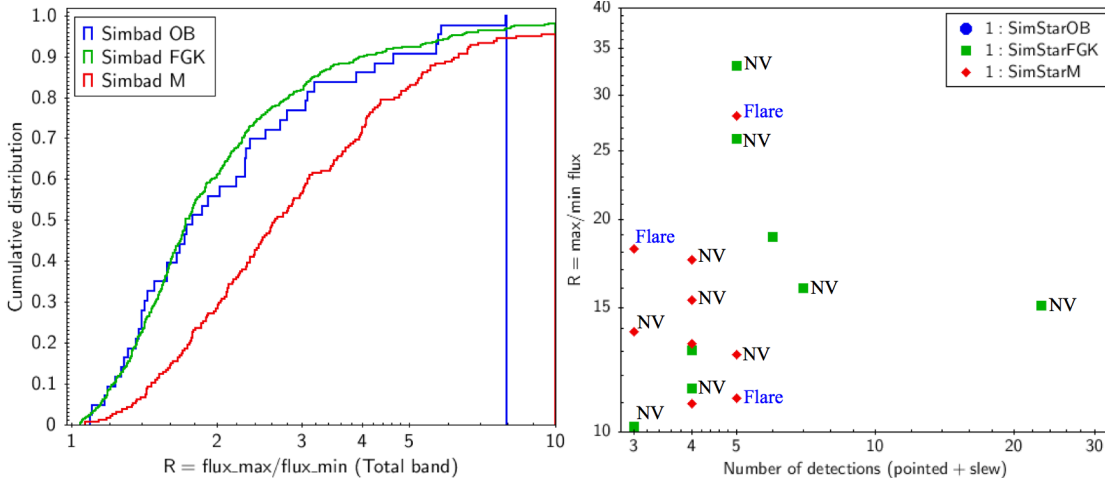


Figure 3.4.3 - Left: Cumulative distribution of $R = F_{\max}/F_{\min}$ ratio (EPIC, total=band 8) for O & B stars (blue), F, G & K stars (green) and M stars (red). Median R values are 1.9, 1.7 and 2.6 respectively, showing M stars are more long-term variable than the earlier classes. Right: F_{\max}/F_{\min} ratio plotted against number of detections in the light curve, for cases with ratios > 10 . Flare labels indicate cases where short-term variability (flares) within observations are seen while NV labels indicate no notable short-term variability within observations.

Some initial science projects were started to explore the LTV data. These include (i) a search for potential tidal disruption event (TDE) candidates that broadly conform to the ‘typical’ $t^{-5/3}$ decay profile, possess a soft spectrum and do not reside in active galaxies, (ii) a comparison of the amplitude of long term X-ray variability in samples of polar and intermediate polar (IPs) magnetic cataclysmic binaries, the distributions lending support to the expectation that IPs are much less variable than polars on long timescales, and (iii) an investigation of long-term X-ray flux changes in cool stars, which shows that later (M-class) stars are more variable than earlier types and that in at least some cases, the variability is not due to short-term flares (Fig. 3.4.3). Preliminary results from these investigations were reported at the EXTrAS science workshop in Pavia, 21-23 November, 2016, and will be pursued beyond the end of the project.

3.5 Screening and validation of results

A standard pipeline processing cannot be expected to reach a 100% reliability of all processed data products and will occasionally produce unreliable results. Tools need to be developed to screen results and products, to make the data products available to the astronomical community in a sensible way and to alert the archive user to potential problems that cannot be overcome during data processing.

The screening tool

To perform screening and validation of data products and results, we developed a tool in the form of a Graphical User Interface (GUI), which allows a quick visualization of EXTraS products from all lines of data analysis (short-term aperiodic variability, pulsations, new transients, long-term variability), as well as of XMM-Newton Processing Pipeline System (PPS) files. The GUI is written in Python and utilizes several modules including matplotlib to display plots and graphics, and pyfits to read and handle FITS files. An interface to the SAOImage ds9 application enables interactive image display. An important feature is to allow the user to generate products, like light curves and images, using different settings (e.g. different energy bands or time intervals) – periodicities can also be independently checked with a Lomb-Scargle periodogram approach.

The vast amount of products generated by EXTraS (more than 20 million files) makes it impossible to inspect them all visually. We devised the following strategy: a sub-sample of observations was selected, to perform a complete visual screening of products from all lines of data analysis, in order to identify and classify issues and problems in the data which can cause false variability of X-ray sources. The outcome of this work was used to improve the algorithms in the processing pipeline as well as to design tests to perform an automatic screening of bulk data products.

Complete visual screening of a sample of results

We selected a sample of 225 XMM-Newton observations. From the analysis of aperiodic variability, we selected only light curves with uniform time binning with optimal bin size and a minimum of 20 data points, and with reduced chi-square larger than 2 for a fit with a constant value (194 sources). A total of 503 sources with periodic signal detected above the confidence threshold of 3.5σ was inspected, some of them being the same object observed at different epochs. We also screened a subset of 189 transient candidates from a preliminary list (produced with the uniform time bin algorithm), and, subsequently, all the 1002 transient candidates from a complete catalogue generated using the Bayesian Block approach. Regarding long-term variability, a problem in the characterization of light curves is related to the way the XMMSAS detection algorithm handles the same object in different XMM-Newton observations – testing the assumptions of a point source or a slightly extended source (in excess to the point spread function). Fluxes of a constant source derived under the two different assumptions are significantly different, resulting in wrong variability parameters. To investigate the effect, a zero-extent (ZE) test version of the XMM-Newton source catalogue was generated, for which the source detection algorithm was restricted to use no source extent. Long-term light curves and images of a sub-sample of 282 sources (selected based on their high variability) were visually screened.

Visual screening of a sub-sample of data showed that most spurious results are caused by only a few effects, prominent in XMM-Newton data. Background variations cause by far most of the spurious periodic signals in periodograms. Since background variations are not strictly periodic, they produce often a forest of signals, which can be used as discriminating parameter. Also several sources in the field of view may show similar periodic signals. Screening of transient candidates has shown a

sizeable fraction of spurious events characterized by short (seconds) timescales. Candidates were classified in four groups with increasing reliability. We investigated the dispersion of the flags allocated by different people to characterize the reliability of the transient candidates. A general agreement was found, which indicates that artefacts among the transient candidates can be recognized by users familiar with the properties of XMM-Newton data. Concerning long-term variability, for a relatively large fraction of sources (after disregarding detections with non-zero extent) reliable variability is found, while most of spurious variability is caused by contributions of various other sources to the X-ray source flux. These comprise instrumental indirect (out-of-time events and single reflections from other X-ray sources within, or near the field of view, respectively) or direct (diffuse X-ray emission or a nearby point source) contamination. Except the case of nearby point sources (the distance to the nearest neighbor is included in the XMM-Newton catalogue) these effects need visual inspection of imaging data which cannot be done automatically.

Screening of bulk data products

Procedures to perform automatic, systematic screening of the bulk data products had been devised in order to identify low-quality (or pathological) cases.

Screening and validation of results from the aperiodic variability analysis been performed at various stages of the work and adopting various approaches. A major bug, that turned out to be related to improper low-level analysis of counting mode time intervals in the PN was identified by systematically comparing light curves taken with different cameras: it has been studied in detail, leading to the decision to run a full reprocessing of PN data. Other bugs were discovered along the way, and some of them have been fixed prior to release; only a few of them, affecting a small minority of products, were left (and will be flagged) in the aperiodic variability catalog. We systematically checked for missing products, through the catalog and the log files: only extremely few cases (<100) are due to problems that could not be addressed. We compared EXTraS light curves and fitted parameters with published results, for example with the light curves of all sources in the rho Ophiuchi star forming region, independently studied within the DROXO project. We run extensive manual screening of products, in particular light curves, checking for anomalies or peculiar behaviors (this strategy revealed spurious blocks in the Bayesian segmentation of bright sources, and motivated the 4 sigma Bayesian block generation). We run MonteCarlo simulation of light curves, that lead to a better understanding and calibration of the spurious blocks occurrence (two columns were added to the aperiodic variability catalog, reporting the expected number of spurious blocks). We statistically checked for outliers in various parameters (this revealed, for example, that often hardness ratios had unrealistically small error bars, motivating a different approach, bootstrap, to their computation).

For sources with periodic signal above threshold, contamination from background or nearby bright sources could in principle be checked by comparing the detected periods for sources in the same observation. Pairs of sources close together contaminate their fluxes which can lead to consistent periods found for both of them. The detection of a similar period from several sources inside the field of view of an observation suggests background contamination. A known spurious signal in the 100s-140s range is related to the counting mode in the pn camera, affecting also periodicity searches on combination of pn and MOS data. However, automatic flagging of different instances of such effects turned out to be unreliable and such information was not included in the catalogue. Visual inspection is needed for the validation of candidate pulsators.

For candidate transient sources, visual screening could be actually performed for all the new sources with $DET_ML > 15$ and duration < 5 ks, leading to the final catalogue containing 136 fully validated transients.

For sources in the LTV catalogue, a number of screening and quality tests have been performed. Firstly, some automated analyses were run to provide independent or semi-independent checks that the various photometric and flag elements of the catalogue are mapped/computed as expected from the input data. To explore the data further, the screening tool has been employed to examine the light curves and the per-observation data (including the images), associated with several samples of LTV sources. This was done at a relatively early stage in the process to identify the more prominent issues, guiding improvements in flagging procedures. It was also performed at the end to confirm that filtering on various quality flags provided in the catalogue removes many previously recognized problem cases, and to try and identify residual situations, for documenting, that give rise to problematic LTV results that are not screened out by use of the quality flag filters. This latter process, for example, helped recognize cases where detection matching may be suspect in complex source regions.

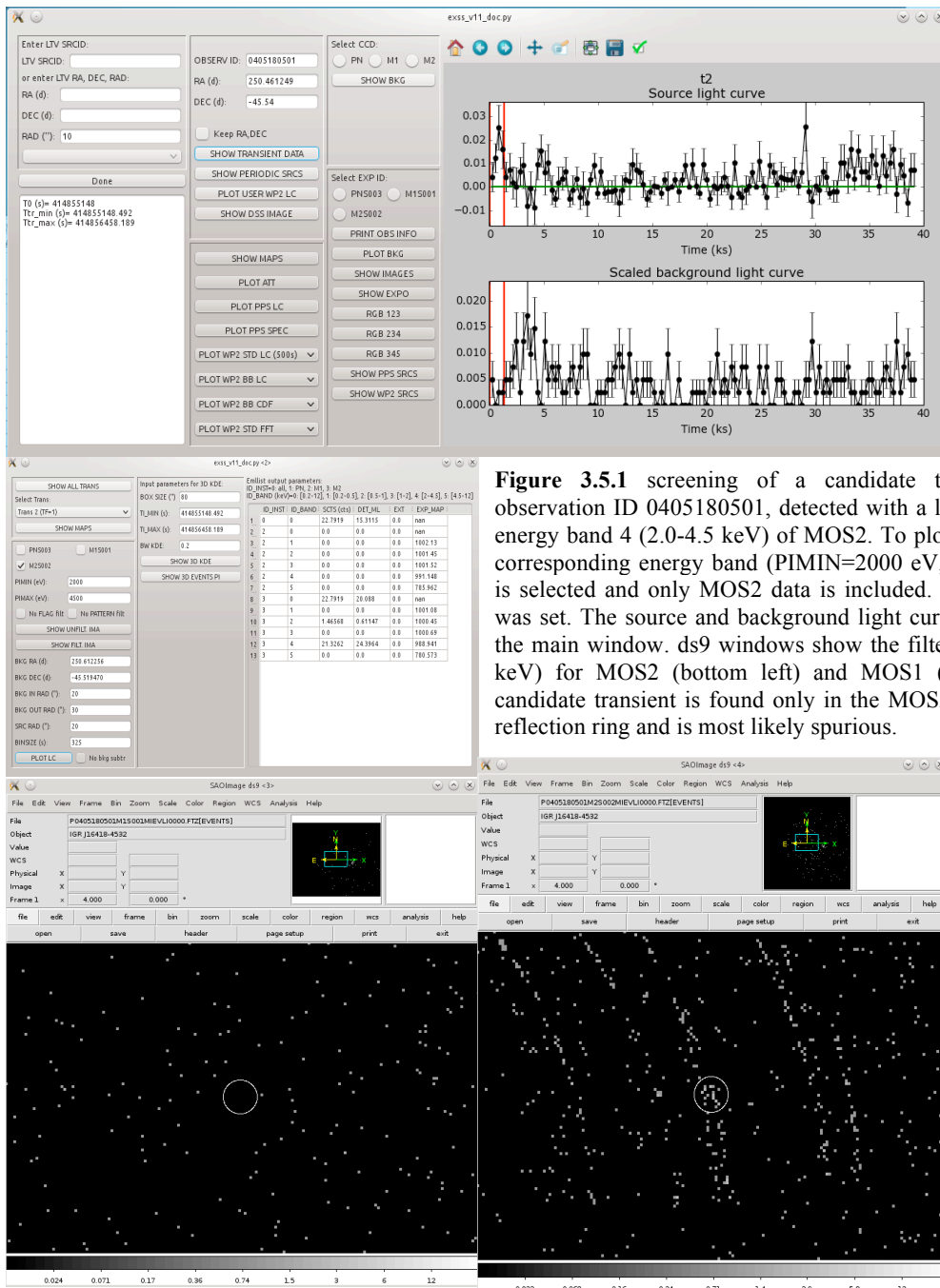


Figure 3.5.1 screening of a candidate transient source in observation ID 0405180501, detected with a likelihood of 24.4 in energy band 4 (2.0-4.5 keV) of MOS2. To plot the light curve the corresponding energy band (PIMIN=2000 eV, PIMAX=4500 eV) is selected and only MOS2 data is included. A bin size of 325 s was set. The source and background light curves are displayed in the main window. ds9 windows show the filtered images (2.0-4.5 keV) for MOS2 (bottom left) and MOS1 (bottom right). The candidate transient is found only in the MOS2 image on a single reflection ring and is most likely spurious.

3.6 Multiwavelength characterisation and phenomenological classification

Originally planned to focus on multi-wavelength characterization of new sources discovered in EXTRaS by the transient search and by the new slew survey, this line was extended to the development of general methods for the classification of X-ray sources, based on the whole set of available XMM sources (as defined by the 3XMM catalogue). This was based on

1. observations of newly discovered sources,
2. catalogue matching of sources with catalogues of sources in other wavebands,
3. use of the variability properties of the sources determined in the project.

The major results are the following:

Source classification

We used both, X-ray temporal and spectroscopic capabilities as well as multiwavelength information.

a. For the characterization of brighter sources (>250 photons) we developed tools to model X-ray spectra. The software extends the standard X-ray astronomical analysis tool ISIS (developed by MIT) by providing capabilities to calculate the statistical uncertainty of best-fit parameters in a parallel environment. This approach significantly speeds up the time needed for the error calculation. The tools were tested by fitting about 137000 spectra from the 3XMM with a set of spectral models. The results of these fits are made available to the community.

b. we developed tools for the characterization of the spectral energy distribution, which allow to treat all multi-wavelength data in their native format without requiring users to flux the data. The fits also treat foreground effects such as the absorption of radiation in the interstellar medium properly. Due to long term source variability, multiwavelength spectra are in principle only representative for a source spectrum if they are simultaneous. For other datasets, the best one can do is to identify periods where a source is not varying "too much", bad coverage in many wavebands makes it difficult to clearly define what this means, and it is clear that any definition of simultaneity will have to include knowledge about the main physical processes probed by a given waveband. As a proof of concept, in an exploratory project (Krauss et al., 2016, A&A 591, A130) we have used a Bayesian blocks decomposition of Fermi lightcurves for a set of 22 radio loud AGN to define phases where the source lightcurves could be described by a constant to derive 81 SEDs that were then modeled using the EXTRaS-developed code.

c. we developed tools that interface to the XMATCH service developed by the EU-funded ARCHES project. The software includes a well defined interface that also handles complex catalogue matches, and works around the limitations of the XMATCH service. For all EXTRaS sources, and the 3XMM, source positions were matched with a set of 53 standard catalogues. Again, to test the tools, an intermediate data release of the 3XMM was used, and multiwavelength information (power law indices between the X-rays and other wavebands) were calculated. 4300 X-ray detections have radio counterparts, 67000 detections have IR counterparts, 54000 detections have optical counterparts, and 39500 detections have counterparts in the gamma-rays. The results of the matches have been made available to the community.

d. In order to classify source, a training sample for machine learning algorithms based on well understood sources was designed. This sample is based on the training sample of Lo et al. (2014, ApJ 786, 20) and Farrell et al. (2015, ApJ 813, 28), but extended by subdividing AGN into Seyfert 1, Seyfert 2, and BL Lac sources based on the newest edition of the Veron AGN catalogue, subdividing, X-ray binaries into High-Mass X-ray Binaries (HMXB) and Low-Mass X-ray binaries

(LMXB) using the catalogues of Galactic X-ray binaries by Liu et al. (2006, A&A 455, 1165 and 2007, A&A 469, 807) and the properties of X-ray binaries in the Small Magellanic Clouds (Haberl & Sturm, 2016, A&A 586, 81), and cataclysmic variables from the Ritter and Kolb catalogue. Using the X-ray spectral and multiwavelength data, and training a random forest classifier, from the spectroscopic information alone the error fraction estimated by this specific classifier is 8.8%, distributed over the individual classes as follows:

	BL	CV	HMXRB	LMXRB	S1	S2	Star	ULX	Class. Err
BL	59	0	0	0	37	4	4	0	0.43269231
CV	0	264	0	0	90	12	30	0	0.33333333
HMXRB	0	0	141	0	3	4	2	0	0.06000000
LMXRB	0	0	0	182	7	2	15	0	0.11650485
S1	0	0	0	1	3690	70	12	0	0.02199841
S2	0	0	1	1	302	708	14	0	0.30994152
Star	0	3	0	0	14	3	1593	0	0.01239926
ULX	0	1	0	0	10	1	9	94	0.18260870

The main confusion is between subtypes of AGN and between CVs and AGN, due to the spectral similarity between these object classes. The sky map of the fully classified 3XMM sources (Figure 3.6.1) shows the celestial distribution of the sources in Galactic coordinates, which is close to what one would expect from the known distributions. The classification will be further improved by incorporating variability information from other WPs. Due to delays in the data delivery, this classification should only become available about three months after the formal end of the project. FAU has identified funding that guarantees the continuation of the classification exercise and its publication through the EXTraS WWW pages.

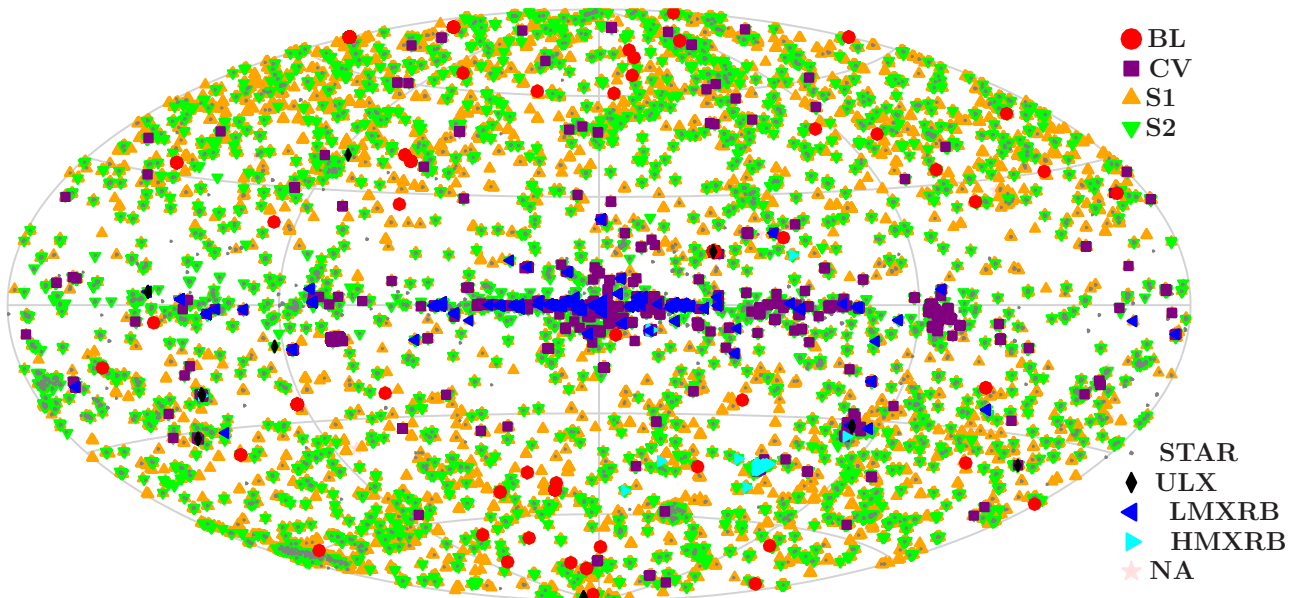


Figure 3.6.1 – Sky map of the classified 3XMM sources (see text for details)

Source follow up

The main results in the area of transient source follow up are the following:

- a. Development of a reduction and analysis pipeline for Swift/XRT data in order to provide a fully automatic routine for transient detection and quick source characterisation. This pipeline was designed for analysis of both daily and archival Swift/XRT data as a training case, and was made modular enough to support XMM-Newton (slew data as well as pointed data) as well. The pipeline also includes a comparison of the source detection list of each new data set to various catalogs (predominantly ROSAT and XMM-Newton) for the identification of long-term transients and variable sources. The pipeline also includes fast source classification algorithms for transient sources with a low number of counts (see above).
- b. With our privileged access to the GROND instrument (mounted at the 2.2m Max-Planck telescope at La Silla, Chile) and the opportunity to observe within minutes (during Chilean night time), we have observed nearly two dozen transients over the EXTraS period (7 XMM slew transients, 14 others). In addition, we also used GROND follow-up observations to optically identify and characterize interesting variable sources discovered by other work packages within the XMM-Newton, primarily pulsars or host galaxies of extragalactic transients. A good fraction of the results were published either in ATels or in scientific papers, with a few more in the process of write-up.

3.7 Archive management

The EXTraS Public Data Archive (EXT-PDA) is an outgrowth of the existing Leicester Database and Archive Service (LEDAS) at the University of Leicester, UK. LEDAS is an online archive for high-energy astrophysics in the UK, and originally grew out of the establishment of the Leicester mirror of the EXOSAT archive in 1989. The Leicester archive now offers immediate access to online archives from most major X-ray satellites, images from the Digitised Sky Survey, and data from over 900 astronomical catalogues. With the inclusion of the largest X-ray source catalogue to date, and its associated data products, the 3XMM-DR4 Serendipitous Source Catalogue (Watson et al. 2013), LEDAS holds in excess of 40 TB of archival data.

The EXTraS Public Data Archive will contain the largest and most systematically generated collection of X-ray photometric data and temporal domain products assembled to date.

The EXTraS Public Data Archive

The EXTraS Public Data Archive (EXT-PDA) was envisioned as a substantial update of the core archive software used by the existing LEDAS ARNIE5 catalogue web interface (<http://www.ledas.ac.uk/arnie5/>). The archive code was completely refactored to modern software standards, using “separation of concerns” between database code and user interface code via a Model-View-Controller architecture. Unit testing was employed for the key database search codes. The update, involving a migration from procedural to object-oriented code structure, is complete. The revised system architecture uses the Flight web micro framework, and the PHP-PDO database abstraction layer making the archive function independent of the choice of underlying DBMS.

The version 1 prototype of EXT-PDA was based on the MySQL DBMS, for ease of transition from the existing LEDAS ARNIE5 system. Following a technical evaluation of competing DBMS software, PostgreSQL was chosen as the core of the final public archive. PostgreSQL enjoys a number of advantages, particularly in the availability of astronomical indexing libraries, and advanced capabilities such as catalogue cross-correlation and ‘footprint’ search within arbitrary geometric shapes.

All 900+ existing astronomical catalogues in the ARNIE5 system have been successfully ported to the new PostgreSQL DBMS, and 2D spatial indexes generated using the H3C sky indexing extension provided by CDS Strasbourg.

Catalogues and bulk products for all primary data analysis lines have been transferred to ULEIC and incorporated in EXT-PDA. A total of 18 TB of LEDAS/EXTraS data is now held on central archival storage at ULEIC. A combined catalogue (“extras_all”) has been generated from the individual catalogues, allowing simultaneous source search across all EXTraS holdings.

The web interface for EXT-PDA has been significantly enhanced. Catalogue search results can be interactively sorted and filtered using the DataTables Javascript extension. Product files associated with a given observation can be selected individually or downloaded in batches as automatically-generated tarfiles.

A visualisation system was designed and implemented via a Python application server running the *Bokeh* graphics library. The system generates interactive time-series graphics in the web browser.

The user command interfaces for EXT-PDA have been refined along the development of the project. The JSON REST interface, which allows basic scripting and returns machine-parseable structured data, is now implemented.

The full EXT-PDA web archive and its interface can be accessed at **archive.extras-fp7.eu**.

As of the end of January, 2017, some work is still ongoing on the archive and its interfaces. Testing and bug fixing is expected to be completed by the end of May, 2017.

Virtual Observatory support in EXT-PDA is currently incomplete – although VOTables can be generated from catalogue search results, full VOSI server support (Graham & Rixon 2011) is still in progress, and will be implemented on a best-efforts basis.

3.8 The EXTraS Science gateway

The software produced in the project is of particular importance for enhancing the potential of discovery of the XMM-Newton mission. This is especially true as we look to the future: the EPIC instrument is still fully operational and collects new data daily, and its operations could be extended for more than one decade. Indeed, the European Space Astronomy Center fully recognized the potential impact of the EXTraS software tools (see <http://www.extras-fp7.eu/images/ESAsupportLetter.pdf>).

Therefore, an important goal we achieved, as stated in the DoW, is the release of several tools developed by our consortium to the scientific community, allowing users to perform customized temporal analysis (aperiodic variability, periodicity searches, transient search) of EPIC data.

The basic, simple solution we adopted is to provide the user with archives, containing all the files required to build each analysis tool. These are released through the EXTraS website.

This approach has been adopted for some important tools, as for the Science Analysis System (SAS). But we are also working to provide most of the software as a set of services through a Web portal – named EXTraS portal, <http://portal.extras-fp7.eu> - designed following the science gateway paradigm. This is a new, important service for the community that was not originally included in the DoW.

The advantages of this approach are clear and straightforward: scientists need only to access the portal, select the observations and the analysis they want to perform, and the portal takes care of all the necessary step to obtain the results.

The architecture of the science gateway is shown in the following Figure. A detailed description can be found at <https://peerj.com/preprints/2519/>.

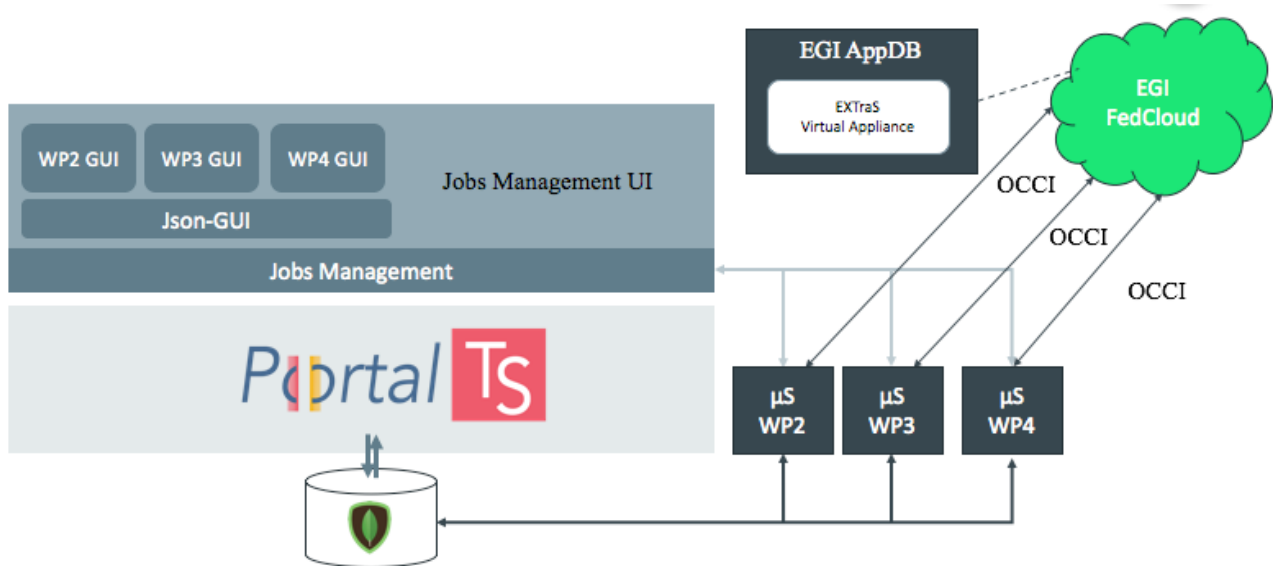


Figure 1 - The architecture of the EXTraS Science Gateway

In order to set up an opportune computational infrastructure, the project signed an agreement (<https://documents.egi.eu/public/ShowDocument?docid=2869>) with EGI, a federated e-Infrastructure partially funded by the European Union (EU) Horizon 2020 program to provide advanced computing services for research and innovation (<https://www.egi.eu/news/egi-and-extras-serving-the-astronomy-and-astrophysics-community/>). The agreement grants the project the possibility to run up to 50 data analysis at a time for one year, but it can be renewed every year.

Presently the EXTraS Science portal supports the analysis tools developed in WP3 (periodicity searches) and WP4 (search for new transients). Installation of WP2 software (aperiodic variability) is currently in progress. As a future development, we plan to upgrade the portal architecture in such a way we will be able to provide our software as a service also for the XMM-Newton pipeline, avoiding to rewrite the code for its insertion into the SAS.

4. Dissemination activities and impact of EXTraS

We summarize the main activities carried out throughout the EXTraS project, with some focus on the project website, on the design of the project logo and avatar, on education activity, as well as on presentations of the project to the scientific community.

4.1 Dissemination activities

Project Website

The final version of the website has been published at the official address: www.extras-fp7.eu on March 2014. It has been designed to address the dissemination of the project results to the scientific community, and to act as a centralized hub for external resources as the EXTraS data archive and the scientific tools for data processing. The web site also includes a restricted area for internal communication. It will have a crucial role also after the end of the project, especially from the delivery of the EXTraS final products.

Although its main objective, we chose to develop also other sections targeted to schools (both teachers and pupils) and general public. The latter one contains a small part addressed to media too.

The sections for schools and the general public are available in English, German and Italian.

The website has been kept up-to-date on a regular basis.

Logo and avatar

An expressive and meaningful logo can be an important tool to diffuse and make recognisable a project with a quick glance. We developed a collaboration with the visual art teacher Alessandra Angelini (Brera Fine Arts Academy) and her students in order to design a logo which had:

- a. to show clearly the name of the project;
- b. to communicate the “variable and transient sky concept”;
- c. to be friendly and welcoming.

The logo (see Annex, fig. A.1) was chosen and used since the very first months of the project.

The visual artist Alessandra Angelini and her students were able to design an avatar too, with similar goals (b. and c., as listed above). The two winners represent a flying-fire young scientist and a flying-fire light source (See Annex, fig A.2) and were officially shown at the Italian Astronomical Society (SAIt) Annual Congress, in Catania, on May the 19th. The avatar will be used to communicate the concept of “variable sky” to children after the end of the formal project.

4.2 EXTraS Education Activity (EXEAt)

Our effort in EXEAt began on January 2015, as soon as the first version of EXTraS software was released. This choice allowed us to assume as a starting point the experience we gathered with the two experimental workshops we held in Milano at the Brera Observatory. We also performed some evaluation analysis which allowed us to fine tune the educational approach. EXEAt is targeted to high school students 16-18 yr old. Its overall goal is to allow trained students to:

- identify variable sources in the EXTraS catalogue using the professional software developed by the EXTraS team;

- classify sources into broad categories through simple parameters;
- understand the general properties of the identified sources;
- understand the unanswered questions related to the identified sources.

Workshops

We run 7 workshops: 4 of them in Italy, 1 in Germany and 2 in the UK. One more workshop will be held in June 2017 at INAF-Brera Observatory, Milano, Italy. We have already selected 24 students in the 17-18 years age range.

In table 1, we summarize the workshop held from the beginning of the project.

Location	No. of Students	Age of Students (y)	Date of workshop	Teachers
INAF-Brera Observatory, Milano, Italy	4	17-18	Feb 2015	S. Sandrelli, A. Belfiore, A. Tiengo, A. De Luca, A. Wolter, G. Trinchieri
INAF-Brera Observatory, Milano, Italy	12	17-18	June 2015	S. Sandrelli, A. Belfiore, A. Tiengo, A. De Luca, A. Wolter, G. Trinchieri
Liceo E. Majorana, Desio, Milano, Italy	30	17-18	Oct 2015	S. Sandrelli, A. Belfiore, A. Tiengo, A. De Luca, M. Canali
Dr. Remeis-Sternwarte Bamberg and MPE, Munich Germany	8	14-17	May 2016	S. Kreykenbohm, I. Kreykenbohm, M. Oertel, A. Irrgang, J. Wilms, H. Hämmerle
INAF-Brera Observatory, Milano, Italy	24	17-18	June 2016	S. Sandrelli, A. Belfiore, A. Tiengo, R. Salvaterra, B. Salmaso, A. Zanutta, A. De Luca, A. Wolter, I. Arosio, M. Carpino, G. Trinchieri
Leicester University, Senior Space School UK	25	7-14	July/Aug. 2016	T. Dickens, N. Tanvir, A. Blain, R. Johnson, M. Wilkinson, S. Rosen
Leicester University, Senior Space School UK	25	7-14	August 2016	T. Dickens, N. Tanvir, A. Blain, R. Johnson, M. Wilkinson, S. Rosen

IT Tools

Since the final version of the EXTraS tools will be at the disposal of the community only at the end of the project – in particular the online version represented by the EXTraS Science gateway (<http://portal.extras-fp7.eu>), we decided to use a virtual machine based on the VirtualBox platform, a free software provided by Oracle, which runs on Windows, OS X and Linux and which can be downloaded from: <https://www.virtualbox.org/wiki/Downloads>

Whenever the VirtualBox platform is installed, it can load the appropriate EXTraS environment with the professional software for data processing reduction, and a large set data extracted from the EXTraS archive. The latter one is prepared by Andrea Tiengo (IUSS), in order to ensure scientific interest in the data reduction performed by the students. It can be downloaded from

An inquiry-based learning activity

We explicitly adopt an Inquiry-based learning strategy, together with a peer-to-peer education approach. This global strategy both foster the critical thinking and engage the students, who act as though they were researchers. A short but clear framework for describing inquiry in astronomy is provided by the IAU project astroEDU and can be found here: <http://astroedu.iau.org/ebi/>

The workshop should be run in a very informal environment. The students should be divided into different groups of 3-4 people.

The first part of the workshop is addressed to engage them in contemporary astrophysics, with a specific regard to the high energy band. Researchers or teachers introduce them to the contents and the technical language of astronomy (as field of view, electromagnetics spectrum, time resolution, space resolution and energy resolution, lightcurve, photons and so on) and maths, (such as probability, confidence level and so on). At this stage, researchers and teachers should try to avoid a trivial “top-down” approach. They are requested to solicit questions and discussion about topics, instead of acting as pure experts. On the other hand, the goal of this part is to give students the right method and conceptual tools to face the data reduction and interpretation, so that some kind of passive learning has to be taken into account by the students.

The second part of the workshop is addressed to allow the students to analyse data from EXTraS. They use the EXTraS software to verify if a candidate source is a real transient. This is made mainly by studying its lightcurve. If the validation is successful, the students should wonder if the X-ray source is a brand new one or an old acquaintance. In the latter case, it should be present in some other X ray archives. They also should wonder if it has a known counterpart in different wavelengths. They can try to match their “new” EXTraS catalogue transient with some other sources in the same position listed in multiwavelength catalogues on-line.

This allows them to “stop-and-try to guess” their transient source.

The leading questions of our inquiry are the following, each of them representing a step for further investigations of the data:

- 1) Is the transient candidate a real transient X-ray source?
- 2) Is it a new X-ray source?
- 3) Has it a possible counterpart at other wavelengths (e.g. optical)?
- 4) What kind of astrophysical object or phenomenon might it be?
- 5) Might it be an important scientific discovery?



School kit

The delivered school kit consists of 5 different chapters.

Chapters 1 to 3 are related to physics background (ch. 1), X-ray astronomy background (ch. 2), XMM-Newton and EXTraS (ch 3). A full set of presentation in English and Italian was prepared. Interested teachers and students can download several lessons and modify them. The modified

versions can be submitted to the EXTraS team, to ensure a major engagement of teachers's professionalism and improve the educational quality of the project itself. Links to other professional

and educational sites are provided.

Chapter 4 is dedicated to EXTraS software. It comprises the necessary links to download the VirtualBox and the set of pre-reduced data extracted from the EXTraS archive. There is also an instruction manual which helps to load the EXTraS environment file on the Virtual Box.

Chapter 5 is a a step-by-step “cookbook”, which guides students and teacher through the whole validation process and the counterpart search. The cookbook is extensively commented, with strong references to the background lessons mentioned above.

Teachers can get also an *EXEAt cookbook for teachers only*, just requesting it at edu@extras-fp7.eu

It contains the five answers for the most interesting transients included in the set of pre-reduced data extracted from the EXTraS archive.

Beyond EXTraS

The presence of a vast catalogue and of the appropriate software – both of them online and freely usable - allows us to think of the EXTraS activity as a Citizen Science candidate project, with a particular interesting activity for in the classroom.

The work-flow could be done entirely online and it could foresee:

- the extraction of transient candidates through the algorithm implemented by EXTraS team
- the validation procedure as described above;
- the counterpart research;
- a final discussion about the results.

4.3 EXTraS Science Workshop

We organized a Scientific Workshop for the community, aimed at reviewing, discussing and exchanging ideas and methods to exploit the rich astrophysical information carried by time variability in the X-rays, for all kind of X-ray emitters - from nearby stars to extreme events at cosmological distances. Synergy of X-ray, multiwavelength and multi-messenger observations, as well as use of current and future serendipitous X-ray data was also discussed.

- Abstracts for 49 contributed talks were submitted to the Scientific Organizing Committee.
- The workshop was held in Pavia at the IUSS premises on November 21-23. It was a success, with 80 attendees from 13 countries. The program included 10 invited speakers and 22 contributed talks, as well as a poster session.
- Early astrophysical results from the EXTraS project were presented. Particular emphasis was put on the potential of EXTraS results and products for addressing different science cases. A special session was also devoted to describe EXTraS analysis and products.
- In coincidence with the workshop, we organized also a public lecture (in Italian) for a broader audience.

For more details, please visit the EXTraS web site.

4.4 Publications

Papers in refereed journals (as of 2017, February)

- *An accreting pulsar with extreme properties drives an ultraluminous x-ray source in NGC 5907*
G.L. Israel, A. Belfiore, L. Stella, P. Esposito, P. Casella, A. De Luca, M. Marelli, A. Papitto, M. Perri, S. Puccetti, G. A. Rodriguez Castillo, D. Salvetti, A. Tiengo, L. Zampieri, D. D'Agostino, J. Greiner, F. Haberl, G. Novara, R. Salvaterra, R. Turolla, M. Watson, J. Wilms, and A. Wolter
2017, *Science*, 355, 817 - arXiv:1609.07375
ESA Press Release: [ESA PR NGC 5907 ULX-1](#)
- *EXTras discovery of two pulsators in the direction of the LMC: a Be/X-ray binary pulsar in the LMC and a candidate double-degenerate polar in the foreground*
F. Haberl, G. L. Israel, G. A. Rodriguez Castillo, G. Vasilopoulos, C. Delvaux, A. De Luca, S. Carpano, P. Esposito, G. Novara, R. Salvaterra, A. Tiengo, D. D'Agostino, and A. Udalski
2017, *Astronomy & Astrophysics*, 598, 69 - arXiv:1610.00904
- *Discovery of a 0.42-s pulsar in the ultraluminous X-ray source NGC 7793 P13*
G.L. Israel, A. Papitto, P. Esposito, L. Stella, L. Zampieri, A. Belfiore, G. A. Rodriguez Castillo, A. De Luca, A. Tiengo, F. Haberl, J. Greiner, R. Salvaterra, S. Sandrelli, and G. Lisini
2017, *Monthly Notices of the Royal Astronomical Society*, 466, L48
- *EXTras discovery of an 1.2-s X-ray pulsar in M 31*
P. Esposito, G. L. Israel, A. Belfiore, G. Novara, L. Sidoli, G. A. Rodriguez Castillo, A. De Luca, A. Tiengo, F. Haberl, R. Salvaterra, A. M. Read, D. Salvetti, S. Sandrelli, M. Marelli, J. Wilms, and D. D'Agostino
2016, *Monthly Notices of the Royal Astronomical Society*, 457, L5
ESA Press Release: [ESA PR M31 PSR](#)
- *Results from DROXO. IV. EXTras discovery of an X-ray flare from the Class I protostar candidate ISO-Oph 85*
D. Pizzocaro, B. Stelzer, R. Paladini, A. Tiengo, G. Lisini, G. Novara, G. Vianello, A. Belfiore, M. Marelli, D. Salvetti, I. Pillitteri, S. Sciortino, D. D'Agostino, F. Haberl, M. Watson, J. Wilms, R. Salvaterra, and A. De Luca
2016, *Astronomy and Astrophysics*, 587, A36

Contributions to conferences (as of 2017, February)

- *A microservice-based portal for X-ray transient and variable sources*,
poster by G. Zereik at the 11th Gateway Computing Environments Conference, 2-3 November 2016, San Diego Supercomputer Center in San Diego, California, USA.
- *Mining the beat of ULXs*,
Talk by G. Israel at Breaking the limits: super-Eddington accretion on compact objects conference, 19-23 September 2016, Arbatax, Italy
- *Hunting in the variable sky - an education activity with XMM-Newton data*,

talk by S. Sandrelli at Global Hands-On Universe (GHO�) Conference 2016 with Galileo Teachers Training Program (GTTP) International Workshop 2016 (22-27 August 2016, Stord/Haugesund University College, Stord, Norway)

- *A microservice-based portal for X-ray transient and variable sources*, talk by G. Zereik at the 8th International Workshop on Science Gateways (IWSG 2016), 8-10 June 2016, Rome, Italy.
- *EXTraS - An experimental didactic program, introducing students to scientific research and analysis using real scientific data*, talk by T. Dickens, at Communicating Astronomy with the Public international congress (May 16-20, 2016, Medellin, Colombia).
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Talk by A. De Luca at *XMM-Newton: The Next Decade* (9-11 May, 2016, ESAC, Madrid).
- *Characterizing the Aperiodic Variability of 3XMM Sources using Bayesian Blocks*
Poster by D. Salvetti, A. De Luca, A. Belfiore, and M. Marelli at *XMM-Newton: The Next Decade* (9-11 May, 2016, ESAC, Madrid).
- *A novel approach to model EPIC variable background*
Poster by M. Marelli, A. De Luca, D. Salvetti, A. Belfiore, and D. Pizzocaro at *XMM-Newton: The Next Decade* (9-11 May, 2016, ESAC, Madrid).
- *Searching for the most elusive X-ray transients with XMM-Newton*, poster by A. Tiengo, G. Novara, G. Lisini, A. De Luca, D. D'Agostino, F. Haberl, J. Wilms, M. Watson, A. Belfiore, R. Salvaterra at *XMM-Newton: The Next Decade*, Conference held on 9-11 May, 2016 at ESAC, Madrid.
- *Unveiling long-term variability in XMM-Newton surveys within the EXTraS project*, poster by S. Rosen, A. Read, D. Law-Green, M. Watson, J. Pye, P. O'Brien at *XMM-Newton: The Next Decade*, Conference held on 9-11 May, 2016 at ESAC, Madrid.
- *Automated classification of new transient sources*, poster by M. Oertel, A. Kreikenbohm, J. Wilms, A. De Luca at *XMM-Newton: The Next Decade* (9-11 May, 2016, ESAC, Madrid).
- *EXTraS discovery of an X-ray flare from the young stellar object ISO-Oph 85*
Poster by D. Pizzocaro, B. Stelzer, R. Paladini, A. Tiengo, G. Lisini, G. Novara, et al. at *XMM-Newton: The Next Decade* (9-11 May, 2016, ESAC, Madrid).
- *SKYLAB: Una finestra sul cosmo*, invited paper by M. Canali (Liceo Majorana, Desio) for the Italian Association of Physics (SIF), 2016, Giornale di Fisica, 3, 225, DOI: [10.1393/gdf/i2016-10250-1](https://doi.org/10.1393/gdf/i2016-10250-1)
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Talk by A. De Luca at *X-ray surveys with advanced multi-wavelength cross-identification methods* (November 30 - December 2, 2015, Paris, France)
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Talk by A. De Luca at *CNOC IX* (22-25 September, 2015, Rome, Italy)
- *The CINEMASCOPE vision of the EXTraS Sky*

Talk by G. Rodriguez at *CNOC IX* (22-25 September, 2015, Rome, Italy)

- *Screening and validation of EXTraS data products*
Poster by S. Carpano, F. Haberl, A. De Luca, A. Tiengo, G. Rodriguez, A. Belfiore, S. Rosen, A. Read, J. Wilms, A. Kreikenbohm, and D. Law-Green at *Exploring the Hot and Energetic Universe: The first scientific conference dedicated to the Athena X-ray observatory*. (8-10 September, 2015, Madrid, Spain).
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Poster by A. De Luca, and EXTraS Collaboration at *Exploring the Hot and Energetic Universe: The first scientific conference dedicated to the Athena X-ray observatory*. (8-10 September, 2015, Madrid, Spain).
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Poster by A. Kreikenbohm, M. Oertel, J. Wilms, A. De Luca, F. Haberl, J. Greiner, C. Delvaux, S. Carpano, D. Law-Green, S. Rosen at XMM-Newton Workshop “*The Extremes of Black Hole Accretion*” (June 8-10, 2015, ESAC, Madrid, Spain)
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Invited talk by A. De Luca at “*Multifrequency behaviour of high energy cosmic sources*” (May 25-30, 2015, Mondello, Italy)
- *Citizien science: ricerca dei buchi neri nel catalogo XMM*
Invited talk by S. Sandrelli at *Italian Astronomical Society (SAIt) Annual Congress* (May 18-22, 2015, Catania, Italy).
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Poster by A. De Luca at “*SWIFT: 10 Years of Discovery*” (December 2-5, 2014, Rome, Italy), *Proceedings of Science*, ed. P. Caraveo, P. D’Avanzo, N. Gehrels, G. Tagliaferri; [PoS online](#); arXiv:1503.01497
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Talk by A. De Luca at *The Universe of Digital Sky Surveys* (November 25-28, 2014, Naples, Italy), *Astrophysics and Space Science Proceedings*, Volume 42, p. 291, ed. by Longo, Napolitano, Marconi, Paolillo, Iodice, Springer International Publishing; arXiv: 1508.07146
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Poster by A. Tiengo presented at *The 40th COSPAR Assembly* (August 2-10, 2014, Moscow, Russia)
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Talk by A. De Luca at *The X-ray Universe 2014* (June 16-19, 2014, Dublin, Ireland)
- *Unveiling long-term variability in XMM-Newton surveys: the EXTraS project*
Poster by S. Rosen et al. at *The X-ray Universe 2014* (June 16-19, 2014, Dublin, Ireland)
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Poster by D. Pizzocaro at *The Unquiet Universe* (June 2-14, 2014, Cefalù, Italy)
- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Presentation by A. Tiengo at *EPIC Calibration and Operation Meeting* (March 25-26, 2014, Garching, Germany)

- *The EXTraS project: Exploring the X-ray Transient and variable Sky*
Poster by A. De Luca at *Explosive Transients: Lighthouses of the Universe* (September 15-20, 2013, Santorini, Greece)

GCN and ATel (as of 2017, February)

- GCN 17548 by A. Belfiore et al.: Swift J0045.2+4151: analysis of XMM-Newton archival data
- ATel #7181 by A. Belfiore et al.: Swift J0045.2+4151: analysis of XMM-Newton archival data

Writings for the general public (as of 2017, February)

- Paper introducing EXTraS (by A. De Luca and A. Simoncelli, in Italian) on “Le Stelle”, astronomy magazine for the general public. August 2015 issue.

4.5 Potential impact of EXTraS

The final EXTraS database describes all kind of variabilities in a sample of hundreds of thousands of soft X-ray sources, on time scales ranging from ~ 0.1 s to ~ 10 years, and on flux ranges spanning from 10^{-9} to 10^{-15} erg cm $^{-2}$ s $^{-1}$ in the 0.2-10 keV energy range. This is the most comprehensive search for, and characterization of variability, on the largest ever sample of soft X-ray sources. A full set of products (light curves, power spectra) is released for each source, together with new software tools. A Science Gateway web service has also been implemented. The impact of our project can be expected to be very broad at different levels:

1. New science will be possible. Variable phenomena are inherent in almost all astrophysical sources, so our catalogue will be a very rich resource for years to come. Several very different ways of exploitation can be anticipated: discovery and characterization of new and unexpected source classes; study of poorly known, exotic phenomena; population studies of different source classes; selection of extreme objects; studies of specific variable sources. The set of early results already published by our team can indeed be considered as the tip of the iceberg of a very large discovery space which is made available to the community.
2. Scientific exploitation of data from XMM-Newton, in itself the most productive observatory of the European Space Agency, will be greatly enhanced. New regions in the parameter space of variable sources will be opened and explored, which would not have been possible through the analysis of single observations. A very broad range of users will benefit from our results. Communities with no expertise in X-ray data analysis will exploit a rich information that would have been otherwise inaccessible, which is a benefit for science per se.
3. A better exploitation of archival data at any wavelength will be stimulated. On the one side, our search can be applied to any astronomical data providing time-resolved photometric information. On the other side, a lot of “follow-up” reanalysis for specific (classes of) sources will be triggered.

4. A solid input for planning future experiments will be provided. A growing interest in time domain astrophysics is permeating the scientific community. Our catalogue will become a reference in the soft X-ray range for years to come, in the era of large surveys. It will serve as a learning case for future experiments focusing on the X-ray variable sky, from SVOM to eROSITA to Athena. On a different level, our project will offer an extensive test for different data analysis approaches and methodologies, which could be directly applied to the analysis of data from future experiments. For instance, our transient search system could be implemented as a (near) real time monitor for transients in future wide-field imaging experiments.
5. The interest of the community in new X-ray experiments for time-domain studies will be triggered, changing the current feeling that in the coming years most of the impetus in transient/variable studies will come from optical or radio telescopes, mostly reflecting an enthusiasm for the “new” (LOFAR, ATA, PTF and LSST), as opposed to the “old” or established (RXTE All-Sky Monitor, Swift, XMM-Newton and Chandra) surveys.
6. The community will be stimulated to promote a better coordination of multiwavelength efforts for transient follow-up studies. Better constraints on known explosive phenomena and the possible discovery of new classes of X-ray transients will allow the assessment of the requirements for a network devoted to multi-wavelength follow-up observations and will contribute to the huge international effort currently in place to identify the gravitational wave and neutrino signal from catastrophic events occurring in the Local Universe.
7. Part of the analysis techniques and tools developed during the EXTraS project could be applied also to the data analysis in different fields beyond astrophysics. For example, within the activities of the IUSS Center for Astronomical and Remote-sensing Observations (ICARO), we have been exploring the possibility to adapt the EXTraS algorithm for the search for transient signals to Earth Observation data.
8. Last, but not least, our project has a great potential for the popularization of science in general and of astronomy in particular. The main focus of our search has a natural appeal, offering excellent opportunities to promote exciting science to a general public audience. As a very specific contribution, our project includes an experimental didactic program that is aimed at directly involving students in our research program, allowing them to participate in the scientific activities. Such a didactic program, which will be interesting per se since it will allow assessment of the potential of a new form of citizen science, could be repeated in the future and its concept could be applied also to other contexts/experiments.

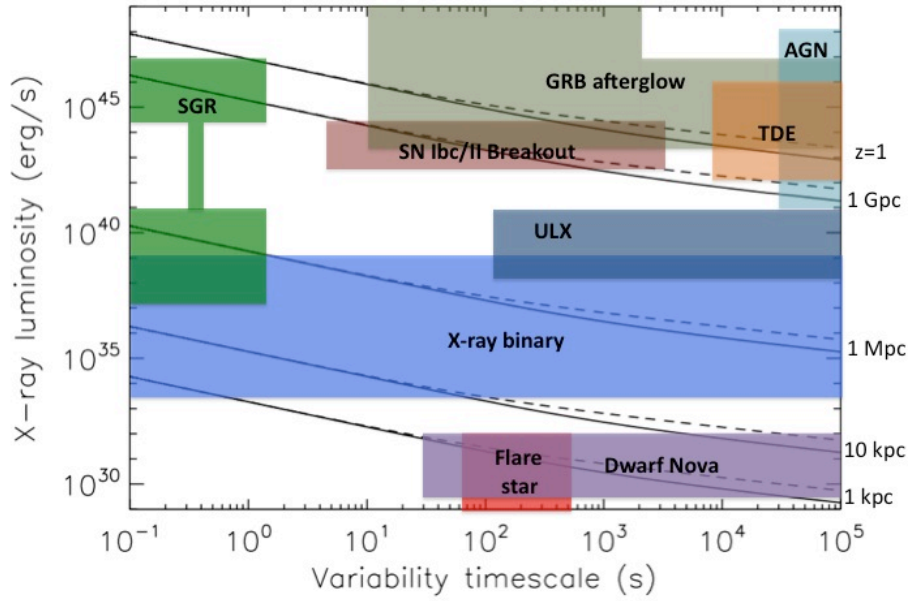


Figure 4.5.1. Peak X-ray luminosity of the most prominent source populations as a function of their characteristic variability time scales (adapted from Merloni et al. 2012). Diagonal tracks indicate the approximate EXTraS 0.5-2 keV 4σ sensitivities (EPIC/pn camera only, see text for details) for various source distances. Solid tracks are computed using the nominal detector background; dashed lines are computed adopting a factor 10 higher background count rate, to show the effect of time periods with enhanced particle background. The extremely large range of variability timescales and luminosities explored by EXTraS ensures that the scientific discovery space is enormous. The possible harvest of variable sources in our catalogue includes a wide diversity of sources, ranging from Galactic objects to the most distant cosmological ones suggesting that the EXTraS results will have a great impact in many areas of astrophysics and cosmology and be of interest for a very broad community. The blind nature of our search will allow us to measure (or give strong limits) on the intrinsic occurrence rate of these sources and to perform population studies. Unexpected, serendipitous discoveries can be also foreseen as the parameter space that is still completely unexplored is extremely large.

Annex

Fig A.1. The EXTraS logo. On the left, the bn version; on the right the full color one, which is usually adopted

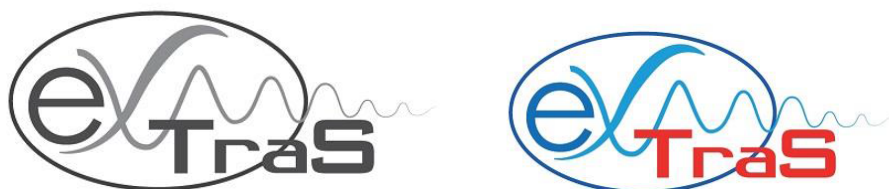


Fig A.2. The avatars. On the left, a flying-fire young scientist; On the right, a flying-fire light source.

