STATiX

https://github.com/ruizca/statix

Detecting X-ray transients in three dimensions

A. Ruiz[1], A. Georgakakis[1], A. Akylas[1], I. Georgantopoulos[1], M. Pierre[2], J.L. Starck[2], M. Chira[1]

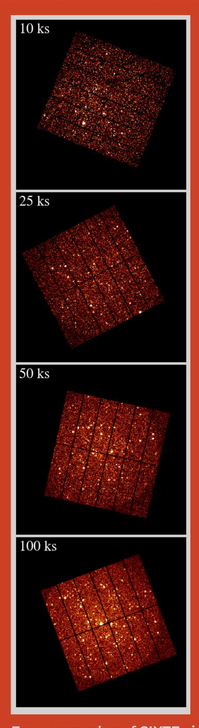
1. IAASARS-NOA
2. CEA-Saclay

IAASARS

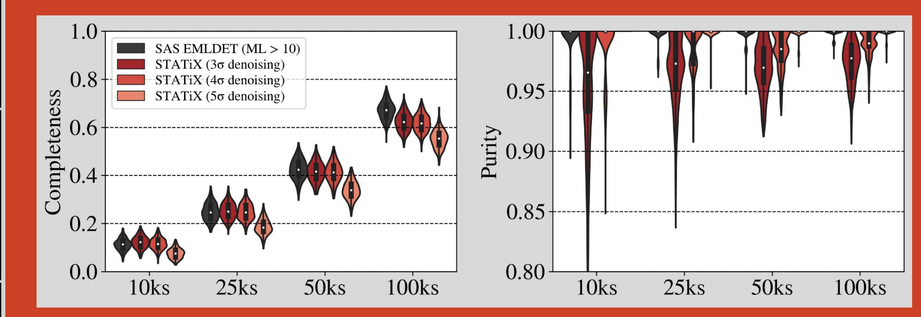
The recent serendipitous discovery of a new population of short duration X-ray transients, thought to be associated with collisions of compact objects or stellar explosions in distant galaxies, has motivated efforts to build up statistical samples by mining X-ray telescope archives. Most searches to date however, do not fully exploit recent developments in the signal and imaging processing research domains to optimise searches for short X-ray flashes.

Here we present **STATiX** *(Space and Time Algorithm for Transients in X-rays), a source detection pipeline that directly operates on 3-dimensional X-ray data cubes consisting of two spatial and one temporal dimension. The algorithm leverages wavelet transforms and the principles of sparsity to denoise X-ray observations and then detect source candidates on the denoised data cubes. The light curves of the detected sources are then characterised using the Bayesian blocks algorithm to identify flaring periods.*

XMM-Newton Simulations

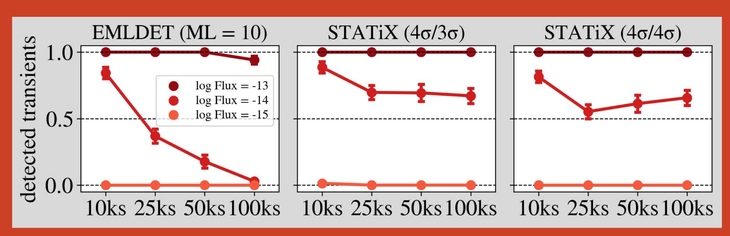


Four examples of SIXTE simulations for the XMM-Newton's EPIC-PN camera. These simulations take into account the geometry of the detector, PSF and vignetting effects, and realistic backgrounds. Each simulation includes a number of constant sources and one 5ks transient. Fluxes of constant sources are drawn from a scale-up version of the logN-logS by Georgakakis et al. 2008.



We used 1000 **realistic simulations of *XMM-Newton* observations to characterise the performance of STATiX**. We used as metrics the **completeness** and **purity** of the resulting source catalogues compared to the input catalogue used for the simulations. The former is defined as the ratio of detected sources and the total number of input sources. The latter is the fraction of true sources among the detected ones. These simulations also allow us to find optimal values for the hyperparameters of our algorithm.

The upper plots show the performance of STATiX for different hyperparameter configurations and exposures times, compared with the EMLDETECT algorithm (2D PSF fitting). The plot below shows the efficiency of detecting flaring sources by EMLDETECT and STATiX for two different sets of hyperparameters, grouped by the flux of the simulated transient. **STATiX can reach levels of completeness and purity comparable to standard detection algorithms while increasing significantly the detection rate of transient sources.**



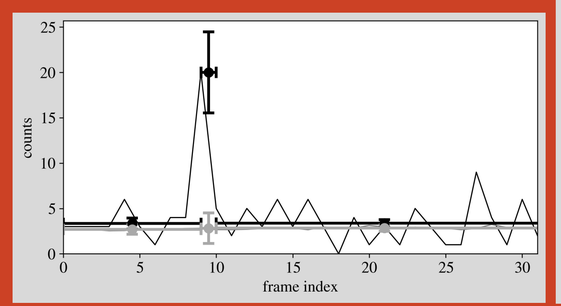
EXTRaS Fields

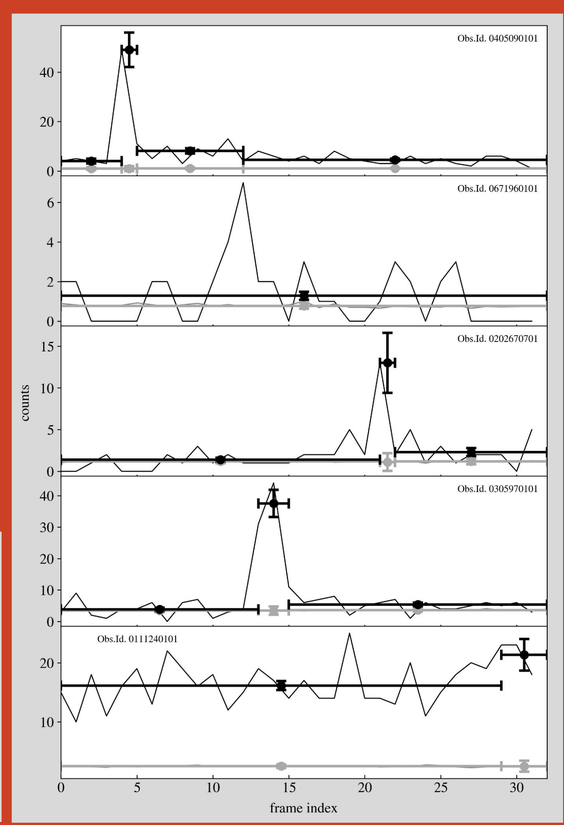
EXTRaS (*De Luca et al. 2021*) is a project aimed at characterising the temporal properties of X-ray sources in *XMM-Newton* archival observation. One of its goals is the identification of short, transient sources.

We selected seven *XMM-Newton* observations containing eight EXTRaS transient candidates suitable for being analyzed with STATiX: observations with clean exposures times above 50 ks and containing a transient source with flaring periods not overlapping any high particle background time interval.

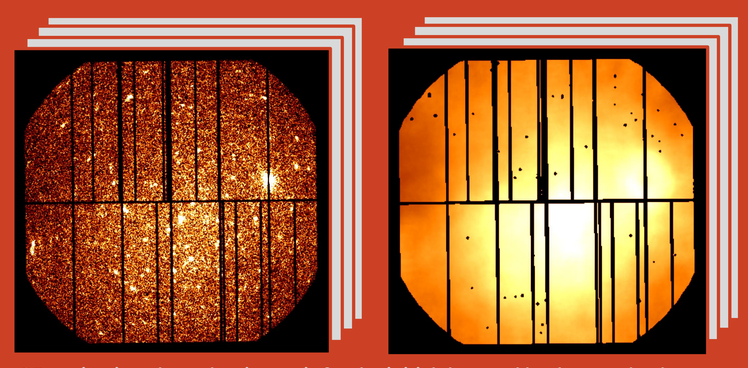
STATiX was able to detect 5 out of 8 EXTRaS transient candidates (~60%) using a denoising threshold of 4σ. The detection rate goes up to 7 out of 8 (~90%) by using a 3σ threshold. We also detected a new transient candidate in these observations not identified by EXTRaS.

The plots show the source (black) and background (gray) light curves for these STATiX detected sources. Solid lines show the raw light-curves extracted from the data-cubes, while the points with error bars show the binned light curve returned by the Bayesian Block algorithm. The horizontal errorbar corresponds to the extent of the time interval. The vertical uncertainty corresponds to the Poisson error.

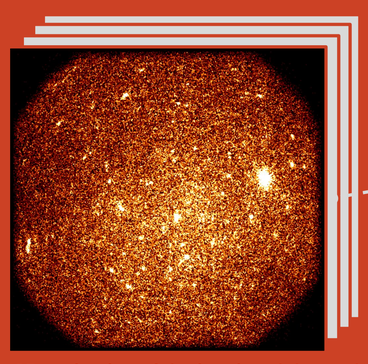




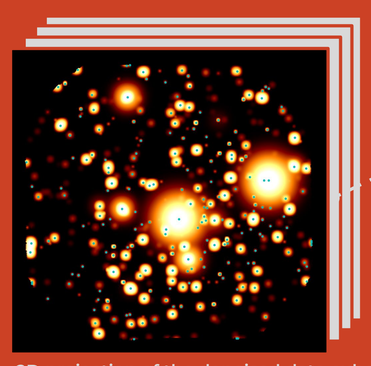
STATiX FLOW CHART



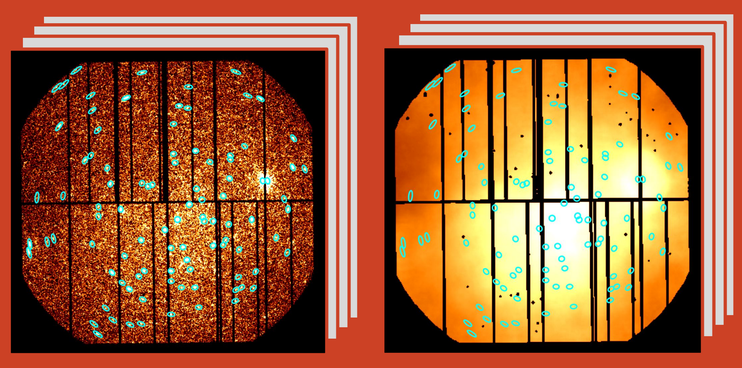
2D projection along the time axis for the initial data and background cubes. Each cube contains 32 time frames. [XMM-Newton Obs.Id. 0305970101]



2D projection of the inpainted data cube.



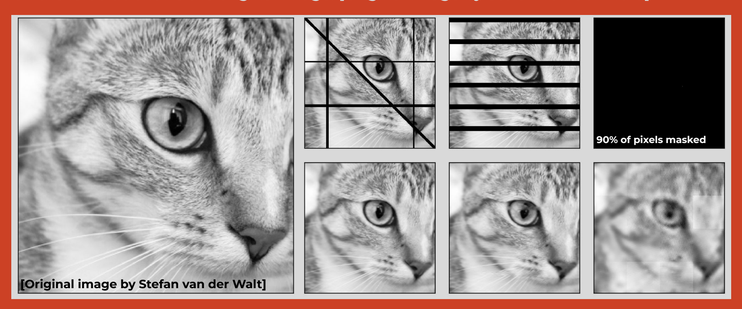
2D projection of the denoised data cube using MSVST. Blue dots mark the source candidates detected with a simple peak detection algorithm.



2D projection of the data and background cubes. Ellipses show the counts extraction regions for statistically significant source candidates after the light-curve analysis.

Inpainting

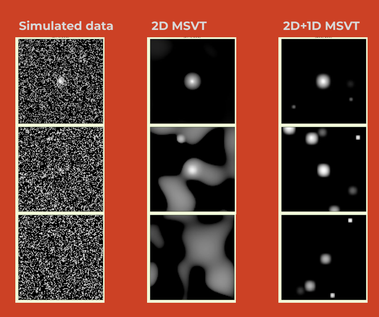
In order to minimise the impact of abrupt changes of the pixel intensity on the source detection algorithm, CCD gaps and bad pixels of the 2D+1D cube are filled using Morphological Component Analysis (MCA, *Elad et al. 2005*), an inpainting algorithm. The figure below shows the performance of MCA for a complex image with texture, prominent edges in horizontal and diagonal directions, as well as features of differing scales. The large panel on the left side shows the original image. [Original image by Stefan van der Walt.]



2D+1D MSVST Denoising

At the core of STATiX lies the Multi-Scale Variance Stabilization Transform (MSVST) presented by *Starck et al. (2009)*. This is a denoising algorithm that attempts to isolate the astrophysical signal in images or data cubes by suppressing the random noise inherent in any observation. This is achieved using discrete wavelet transforms to decompose images or data cubes into a set of wavelet functions with different scale parameters.

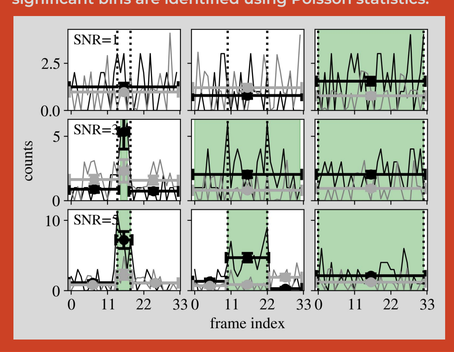
In STATiX we use the 2D+1D version of MSVST, where the spatial and temporal dimensions are assumed to be independent and are analysed separately by defining wavelets that can be expressed as the product of one spatial (2D) and one temporal (1D) component. As a result the original data cube can be represented by a set of wavelet coefficients that correspond to different combinations of spatial and temporal scales. For each of these coefficients a VST (Zhang et al. 2008) is applied, transforming the noise distribution into gaussians of known standard deviations (σ). Noise filtering is then calculated via gaussian thresholding, setting to zero the wavelet coefficients below a certain threshold derived from the VST σ values. The final denoised data cube is reconstructed through an iterative process using the filtered wavelet decomposition.



Example of application of MSVST denoising to simulated data. The panels on the left show images containing a single source (different fluxes for each row) and Poisson noise (same background level in all rows). The central column shows the results of the 2D version of MSVST, while the panels in the right column show the output of the 2D+1D MSVST algorithm. The later is significantly more efficient in reducing the noise level, showing less extended artifacts and helping in the detection of the faintest source.

Bayesian Blocks

Counts are extracted at the positions of source candidates from the individual frames of both the original data cube and the corresponding background map. This step yields both source and background light curves with a time resolution determined by the number of cube frames. Light curves are analyzed using the Bayesian Blocks (BB) algorithm (*Scargle et al. 2013*) to find the optimal binning for the count series. Sources with approximately constant flux in time are expected to be assigned a single time bin. The light curves of flaring events are expected to be broken down into multiple segments. Statistically significant bins are identified using Poisson statistics.



Each panel shows the source (black) and background (grey) light curves for transient objects with different signal-to-noise ratios, and different flaring intervals (dotted vertical lines). The circles show the bins returned by BB. The green shaded regions mark BB segments for which the total number of counts given the background level is > 3σ statistically significant.