XPhotonBank

Photon-level data organization for flexible X-ray analysis

Zhen Zhang[1,2], Yunfei Xu[1,2], Yuan Liu[1], Dongyue Li[1], Huaqing Cheng[1], Chenzhou Cui[1,2], Weimin Yuan[1]

[1] National Astronomical Observatories, CAS

[2] National Astronomical Data Center

01 Why

For long-term analysis and new source discovery

EP's observation-based pipeline bins photon events into images before source detection. As data accumulate, studying the long-term evolution of light curves and energy spectra of variable sources will become increasingly important. However, observation-based data distribution and analysis require downloading individual observation files. The large field of view and frequent observations make this particularly demanding (Figure 1A). A fixed exposure also limits discovery: faint sources need more accumulated exposure, while short bursts can be diluted by background (Figure 1B). Photon-level access allows the analysis region and time scale to follow the scientific question.

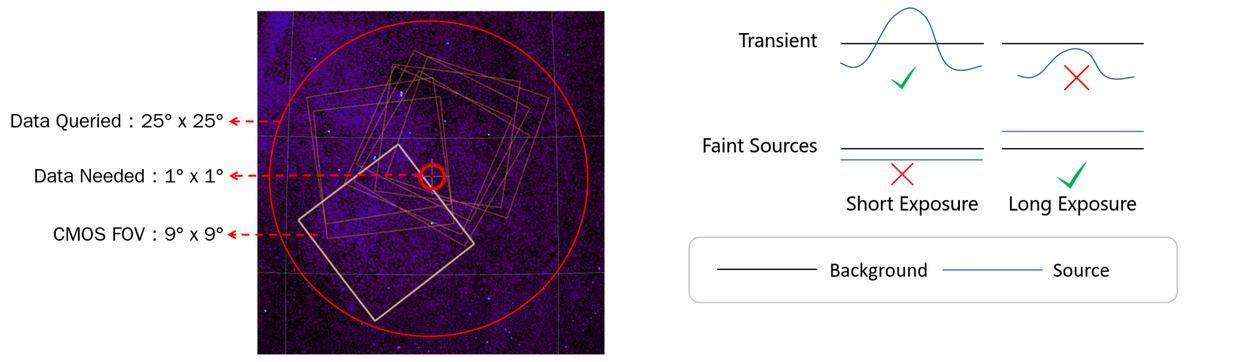


Figure 1. The necessity of photon-level data analysis. **(A)** Due to the ultra-large field of view of EP, acquiring long-term data at a specific location requires downloading hundreds of times more data than necessary. **(B)** Fixed exposure time leads to many sources being missed. Faint sources need more accumulated exposure, while short bursts can be diluted by background.

02 What

A reorganized photon-event database and supporting toolkit

XPhotonBank reorganizes calibrated EP events in RA–Dec–time–PI space (Figure 2). HEALPix indexing and a time-series database support photon queries, with good time intervals stored for exposure calculations.

The accompanying tools support:

- High-dimensional data visualization
- Image stacking
- Photon-level long-term light curves
- Dynamic sky movies
- Clustering analysis for transient discovery

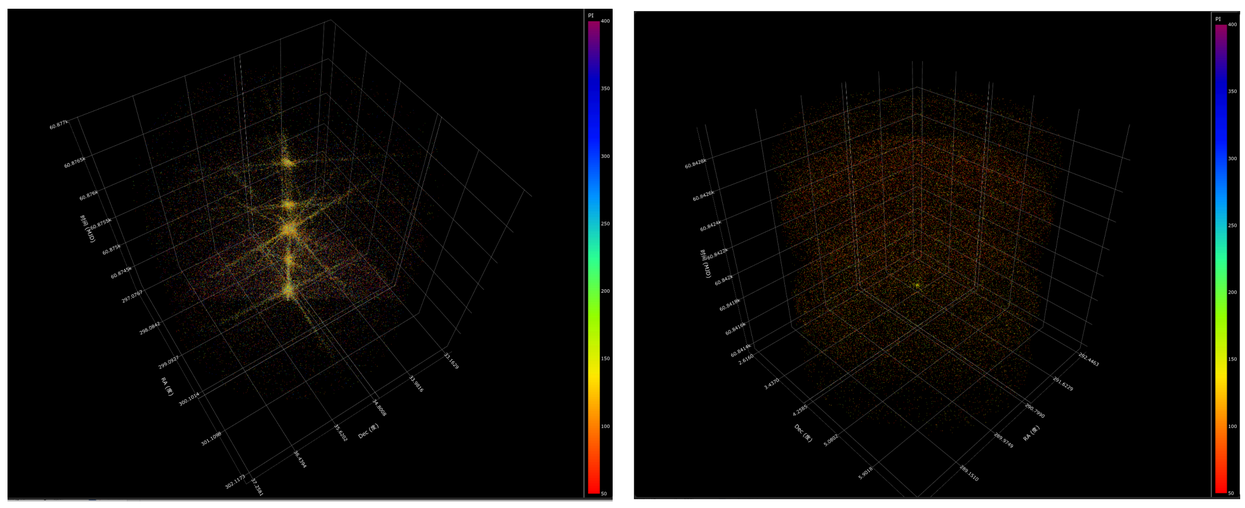


Figure 2. Photon space–time-energy distributions. **(A)** A persistent source with its rotating lobster-eye point-spread-function arms. **(B)** A transient shows temporal concentration and spectral evolution. Axes: *RA, Dec, time*; color: PI (energy channel).

03 Application

Nearly an order of magnitude better sensitivity to faint sources and short transients

To address the two goals above—faint-source and transient discovery—we developed data-stacking and density-based clustering algorithms using the photon database. Several transients, including EP251020a and EP J012607.7+121047, have been reported in The Astronomer's Telegram (ATel). We quantitatively evaluated the performance of both algorithms. For faint sources, we plotted flux histograms of sources detected by the standard pipeline and photon-level stacking (Figure 3A). The shift in peak values between blue (stacked detection) and orange (standard detection) demonstrates a **8 times improvement** (from 3.0×10^{-11} to 3.9×10^{-12}) in sensitivity. For transients, we injected simulated bursts with different fluxes and durations into the photon database, then attempted to recover them using the clustering algorithm and the standard pipeline. For a given burst duration, the results (Figure 3B) show a sensitivity **improvement of 6 times** (the envelope line).

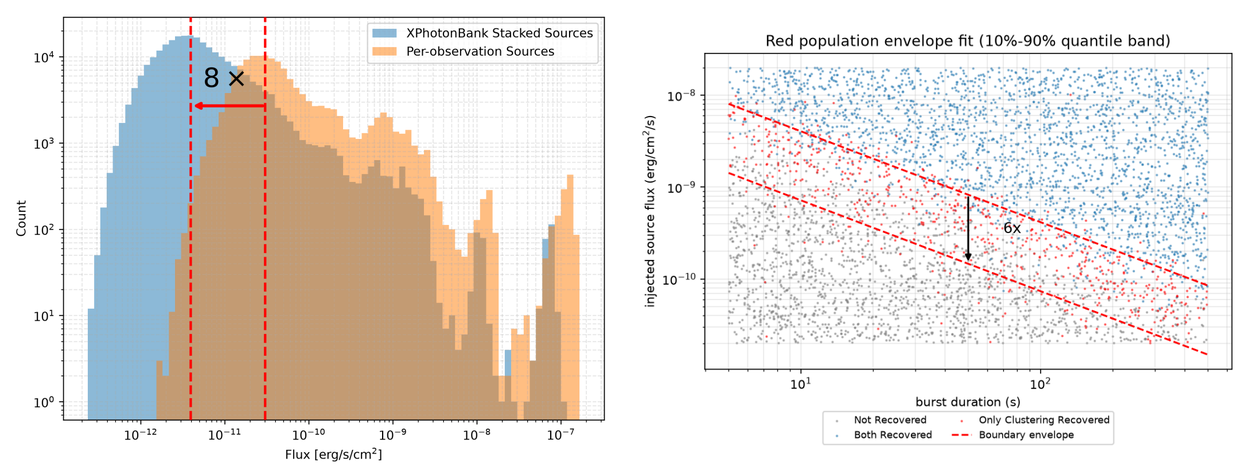


Figure 3. Stacking and clustering performance. **(A)** The shift in peak values between blue (stacked detection) and orange (standard detection) demonstrates a one-order-of-magnitude improvement (from 3.0e-11 to 3.9e-12) in sensitivity. **(B)** Scatter for the recovery of simulated transient sources. The clustering algorithm (red) recovers bursts at lower fluxes than the standard pipeline (blue), with a sensitivity improvement of 6 times (the envelope line).

04 Future

More photon-level algorithms and high-energy missions

Although transients such as EP251020a have already been identified, we expect the sensitivity can be further improved by more fine-tuning. In high-dimensional space, there are also various methods for detecting and identifying transients that are worth studying. Finally, the database architecture and algorithms can be adapted to other high-energy missions, such as eXTP and eROSITA, with instrument-specific adjustments.

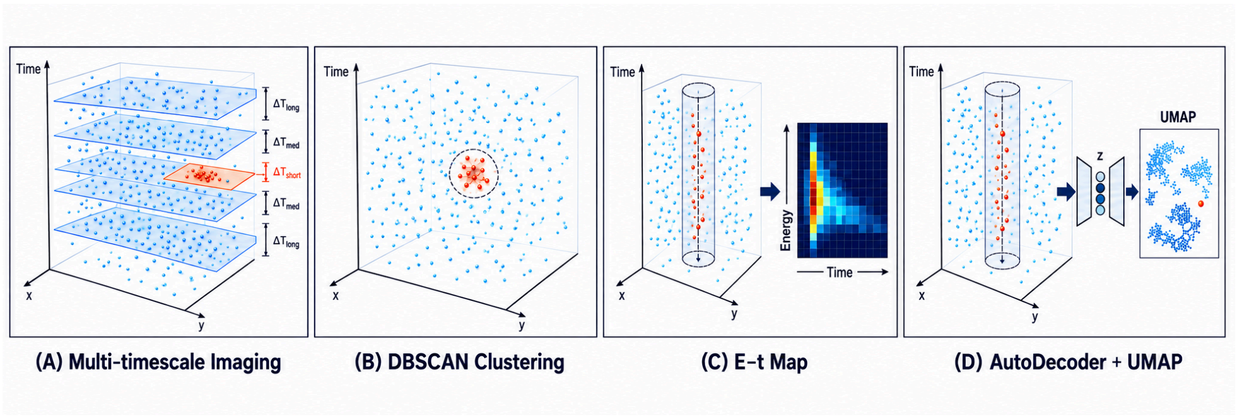


Figure 4. Possible research directions. **(A)** Stacking and slicing at different time scales. **(B)** Density-based clustering. **(C)** Joint time–energy feature analysis. **(D)** Encoding photon sequences as high-dimensional vectors.

Contact Me: zhangzhen@nao.cas.cn