Reprocessing all the XMM-Newton scientific data: a challenge for the Pipeline Processing System

José Vicente Perea-Calderón, Pedro Rodriguez-Pascual, Carlos Gabriel

XMM-Newton Science Operation Center. European Space Astronomy Center (ESAC – ESA)

2019 will mark the 20-year anniversary of the XMM-Newton Mission. So far, the mission has successfully completed a total of around 14.000 pointing observations, and it is expected to continue for many more years, producing a huge number of high-quality science data products.

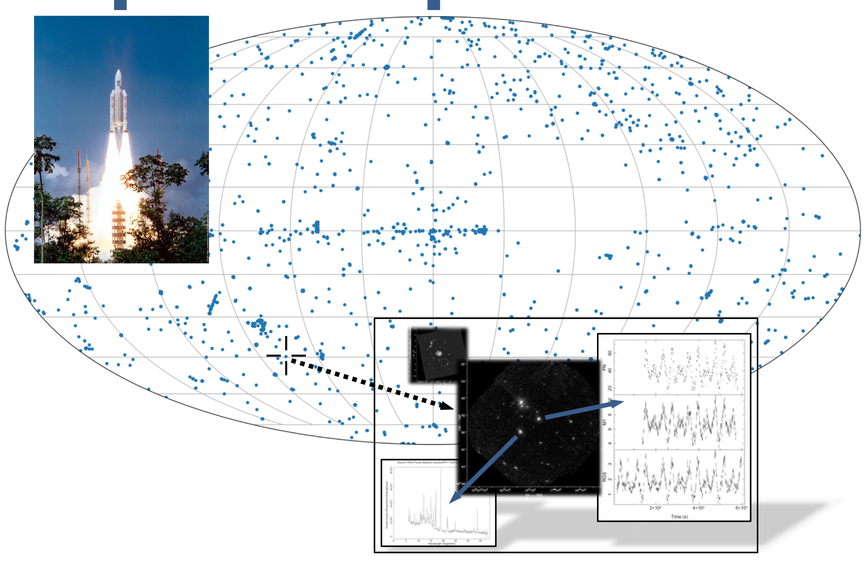
Data processing of those observations is carried out by the XMM-Newton Pipeline Processing System (PPS) and the products are delivered to the XMM-Newton Science Archive (XSA). During the past two decades many changes have been implemented in the data processing software, partly following improvements to the calibration of the science instruments. Several re-processing campaigns have been undertaken along the mission in order to have an up-to-date and uniformly processed set of high-level science data products in the archive. A new re-processing exercise is to be carried out, as it has been more than six years since the last re-processing campaign in 2011.

Unlike the daily mission operations where a limited number of observations have to be processed by PPS, a whole mission re-processing is a real challenge. An individual XMM-Newton Pipeline job (of one observation) can take up to six hours of computer processing time, some of them even longer. To achieve the processing of thousands of observations in a reasonable period of time requires a special preparation including a deep analysis of the computing resources. An extreme optimization of the resources sharing becomes essential in our case.

Besides the optimization of the computing infrastructure usage, a set of software tools had to be developed in order to cope with the management and monitoring of this enormous number of individual Pipeline jobs.

A long life mission ➔ bulk re-processing campaigns needed to populate the archive with the best up-to-date high-level science data products

1999 — 2003 — 20 years of mission — 2019

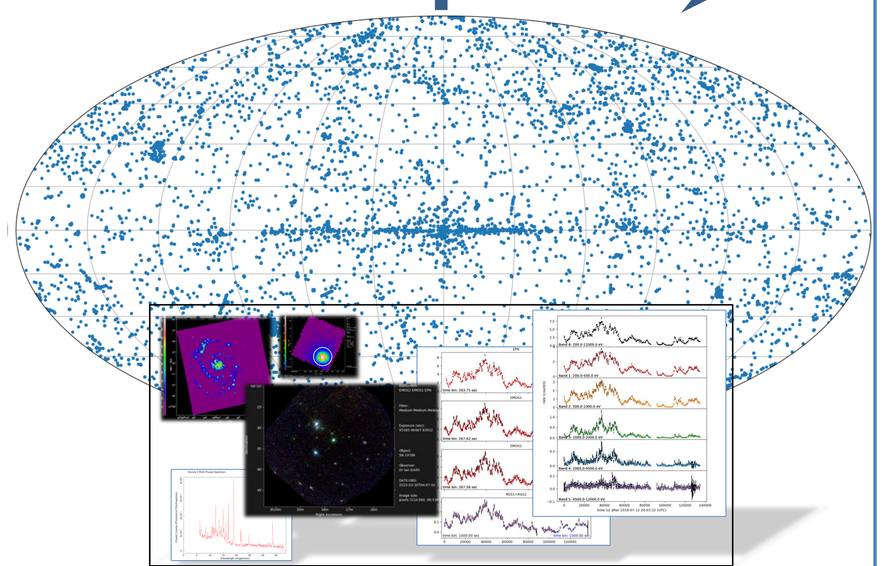


Observations number ↗

Calibration complexity and associated software ↗

Time domain exploration

Improved and brand-new products



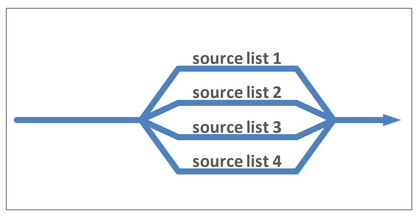
The XMM-Newton Pipeline has improved the quality and quantity of the resulting data products over the years. These products are provided for the main target and also for every source detected in the field. Products include spectra, time series in different energy bands, background estimate, … in different visualization formats.

14.000 observations !!!

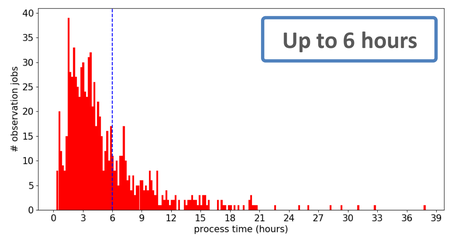
The challenge of processing the whole mission in a short period of time (weekly ?)

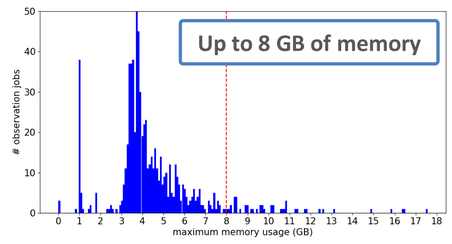
XMM-Newton Pipeline single-job analysis

The first approach to speed up the process is splitting every Pipeline job into 4 threads. Process time reduction by a factor 2.



But still, most of the Pipeline jobs take less than 6 hours to complete in any of the computer nodes of the ESAC/ESA Grid infrastructure. In addition, many of those jobs might need up to 8 GB of memory to be processed:





XMM-Newton Processing computer infrastructure

ESA - European Space Astronomy Center (ESAC)

45 computer nodes
48 – 256 GB
16 – 48 cores

XMM-Newton Pipeline Processing System (PPS) - PRESENT –

conservative use of resources

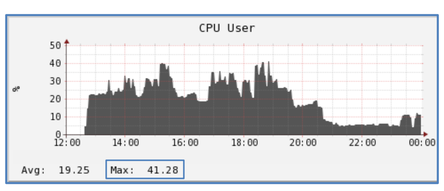
277 simultaneous jobs

1 job: 4 cores (one thread per CPU core) and 8 GB of memory

Conservatively so far, the Pipeline has been initially set up to request a limited amount of memory and CPU resources

Following a testing phase of high workload experiments on the computer infrastructure …

The **maximum memory demand** of the jobs only happens in very short time peaks, so there is a lot of free memory most of the time. **CPU's load** hardly reach the 50 % of the total infrastructure CPU power



140%

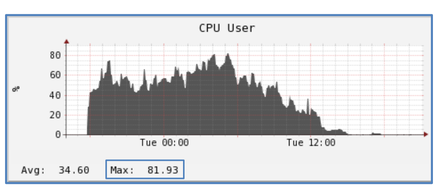
XMM-Newton Pipeline Processing System (PPS) - FUTURE –

high increase of sharing resources

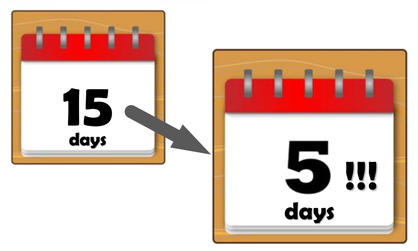
668 simultaneous jobs

1 job: 2 cores (2 threads per CPU core) and No memory constraint

We can further intensify sharing of the memory and CPU resources in the computer infrastructure:



All mission re-processing



– Conclusions –

- We can reprocess all the mission at any time !
- Requirements for every single-job matter
- Deep analysis of the computer capabilities is essential to accomplish the process of this huge number of jobs within a reasonable time
- New monitoring and management tools become a necessity