Mathematics & Statistics | Lancaster University

Efficient Online Detection of Changes in Data Streams: The FOCuS Algorithm

Paul Fearnhead and Giuseppe Dilillo
Lancaster University and INAF-IAPS

Joint work with Gaetano Romano, Guillem Rigaill, Kes Ward and Idris Eckley

Online Change-Point Detection

Online Change Detection

Increasingly there is the need to detect changes in the features of (a) data stream(s) in real-time.

This involves repeatedly testing for a change as each new data point arrives.

We will consider detecting a change in the mean of a single data stream.

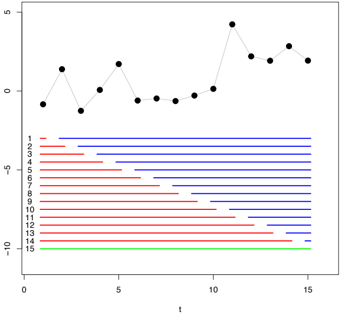
Likelihood Ratio Test

Mathematics & Statistics | Lancaster University

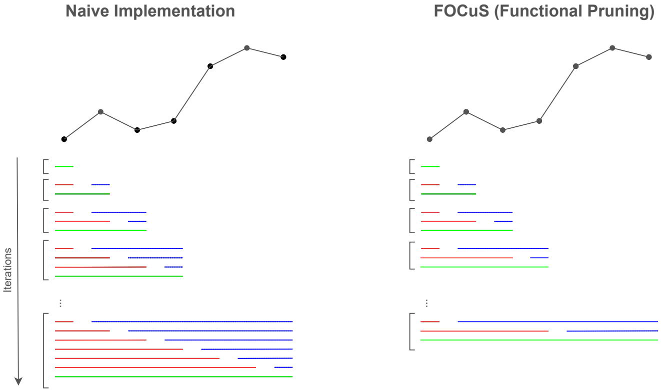
We can test for a change by repeatedly using a **likelihood ratio test**.

We compare evidence for a model with no change against a model with a different mean before and after a changepoint τ.

As the **changepoint is unknown**, we search over all possible τs.



The Idea of FOCuS



Mathematics & Statistics | Lancaster University

FOCuS Algorithm

Page's method

Consider Gaussian data, with known pre-change mean (w.l.o.g. assume this is 0). Page (1954) introduced a sequential change detection method – it requires you to specify the post-change mean, μ.

It calculates the maximum likelihood ratio statistic for a change to a mean μ, maximising over all possible change point times.

$$P_t(\mu) = \max_{\tau=1,\dots,t} \sum_{s=\tau}^{t} \{y_s^2 - (y_s - \mu)^2\}.$$

This can be done online using the recursion

$$P_t(\mu) = \max\{P_{t-1}(\mu), 0\} + \{y_t^2 - (y_t - \mu)^2\}.$$

FOCuS Algorithm

We do not know the post-change mean μ. So ideally we would run Page's method simultaneously for all μ.

If we can do this, then $\max_\mu P_t(\mu)$ is the likelihood-ratio statistic. (Equivalent to maximising over all possible change locations.)

This is what the Functional Online CuSum (FOCuS) algorithm does.

FOCuS Algorithm

The idea is to solve

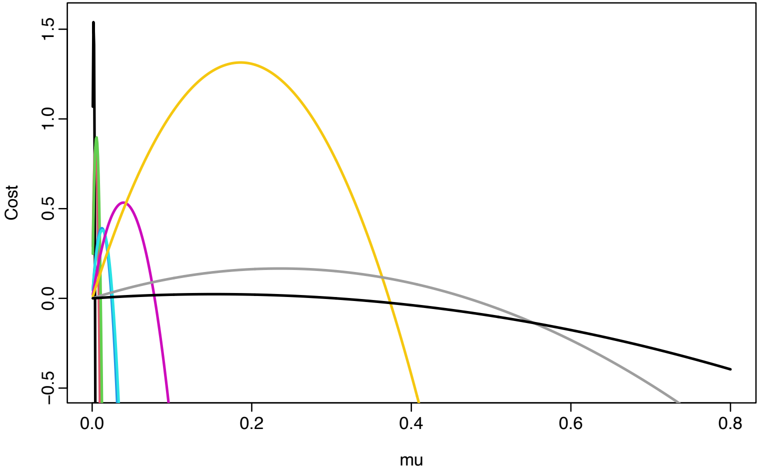
$$P_t(\mu) = \max\{P_{t-1}(\mu), 0\} + \{y_t^2 - (y_t - \mu)^2\},$$

for the functions $P_t(\mu)$ for $t = 1, \ldots$. Each function $P_t(\mu)$ can be written as a piecewise quadratic.

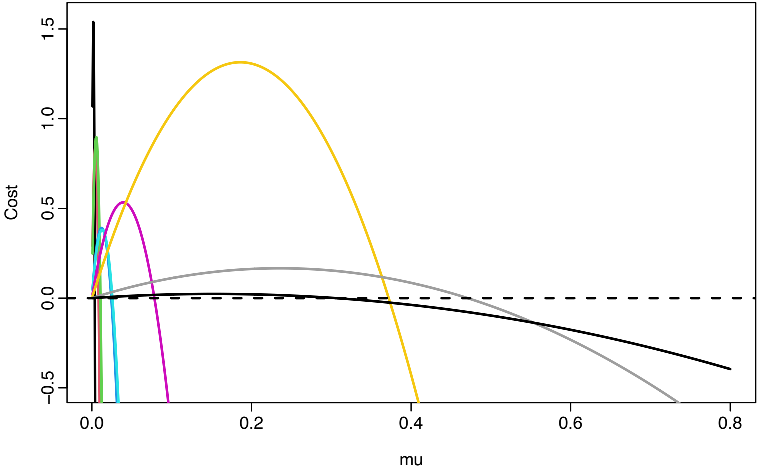
We will describe how to solve the recursion for $\mu > 0$ (and by symmetry the same approach can be use for $\mu < 0$).

FOCuS Algorithm

P_{t-1}

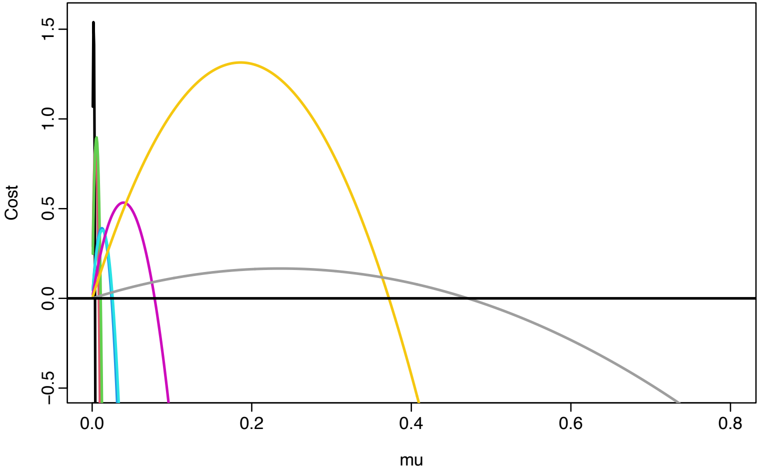


FOCuS Algorithm
Add Zero line



FOCuS Algorithm
$\max\{P_{t-1}, 0\}$

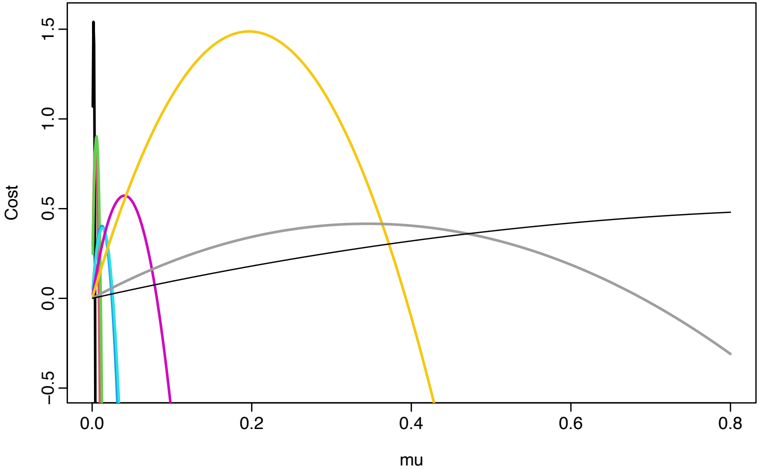
Mathematics & Statistics | Lancaster University



FOCuS Algorithm

$$P_t = \max\{P_{t-1}, 0\} + \{y_t^2 - (y_t - \mu)^2\}$$

Mathematics & Statistics | Lancaster University



FOCuS Algorithm CPU Cost

Each quadratic goes through the origin, so can be stored as a pair: the coefficients of μ^2 and μ, together with the region of μ that it is optimal.

We can show that at each iteration the computational cost is proportional to one plus the number of quadratics that are removed.

As only one quadratic is added at each iteration – this means the computational cost of updating the quadratics is, on average, constant over time.

In practice it is negligible compared to the cost of maximising $P_t(\mu)$.

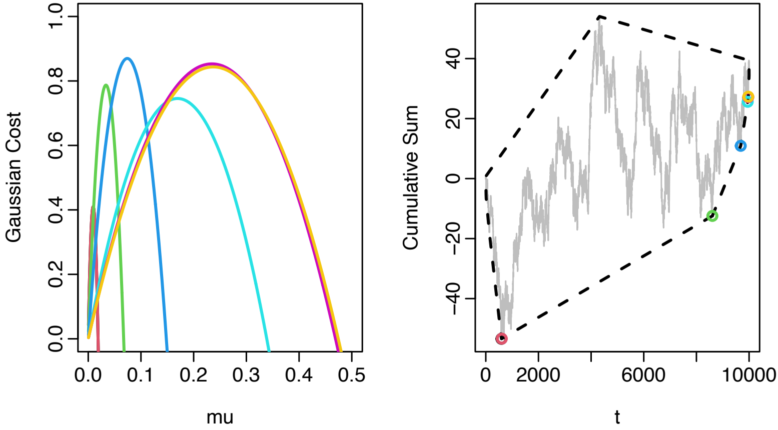
FOCuS Algorithm CPU Cost

There is an additional cost of calculating $\max_\mu P_t(\mu)$ at each iteration.

This involves maximising each quadratic.

The number of quadratics is equal to the number of extremal points of the convex hull of the random walk $(t, \sum_{i=1}^{t} x_t)$.

FOCuS Algorithm
CPU cost



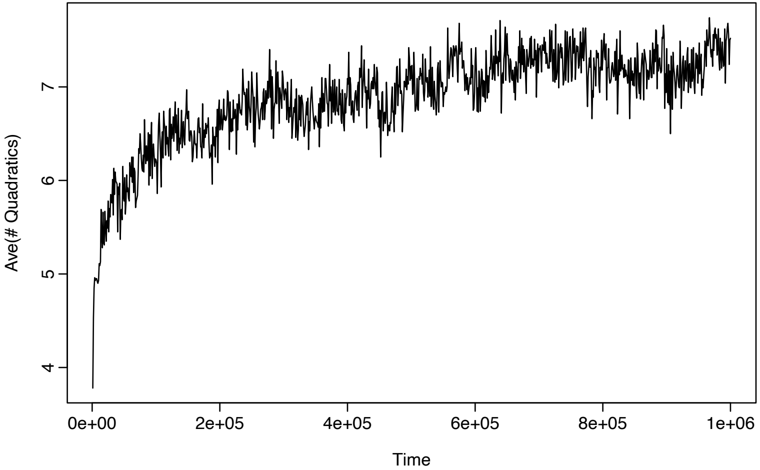
FOCuS Algorithm
CPU Cost

The expected number of points on the convex hull for exchangeable data is well known.

At time T the expected number is bounded by $1 + \log T$.
(Actually $1 + \frac{1}{2} \log T$.)

FOCuS Algorithm
Number of Quadratics

Mathematics & Statistics | Lancaster University



Extensions

FOCuS Algorithm
Unknown pre-change mean

If we do not know the pre-change mean the LR test is:

$$\max_{\substack{\tau\in\{1,\dots,n-1\}\\ \mu_0,\mu_1\in\mathbb{R}}} \left\{\sum_{t=1}^{\tau}(x_t-\mu_0)^2+\sum_{t=\tau+1}^{n}(x_t-\mu_1)^2\right\}-\max_{\mu}\sum_{t=1}^{n}(x_t-\mu)^2.$$

This is easy to calculate given τ, but involves a maximisation over τ.

However there is a small set of τ values we need to consider. These correspond to the values that need to be kept if the pre-change mean is known, for some value of μ_0.

For a positive (negative) change this set is the set of possible change locations in the limit of μ_0 going to $-\infty$ (respectively, ∞)

FOCuS Algorithm
Unknown pre-change mean

This means we can use the same pruning algorithm for the pre-change mean known case. (With some minor tweaks.)

The expected number of quadratics at time T is bounded by $1 + \log(T)$ for each of an up and a down change.

FOCuS Algorithm
Exponential Family

What if our data is not Gaussian? Instead assume it is from a one-parameter exponential family model

$$f(x \mid \theta) = \exp\left[\, \alpha(\theta) \cdot \gamma(x) - \beta(\theta) + \delta(x) \,\right].$$

Page's methods for calculating the LR test statistic for a change from θ_0 to θ has recursion

$$P_t(\theta) = \max\{0, P_{t-1}(\theta)\} + 2\left\{[\alpha(\theta) - \alpha(\theta_0)]\gamma(x_t) - [\beta(\theta) - \beta(\theta_0)]\right\}.$$

FOCuS Algorithm
Exponential Family

Mathematics & Statistics | Lancaster University

We could apply the same type of algorithm to compute $P_t(\theta)$ but with a different form for the curves. These will be of the form

$$a \log \theta + b\theta$$

for some co-efficient a and b for Poisson and Gamma models, or

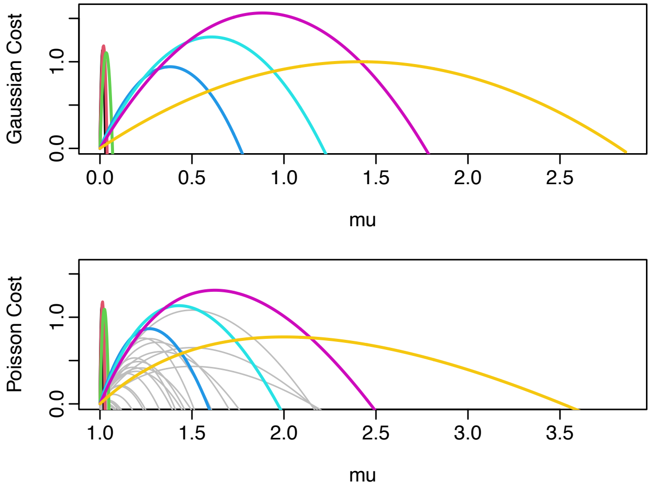
$$a \log \theta + b \log(1 - \theta)$$

for Binomial data.

We can show that the set of τ values we need to consider are the same as for FOCuS for Gaussian data applied to data $\gamma(x_t)$.

The only thing that changes is the form of the curves and hence the maximisation step.

FOCuS Algorithm
Poisson Data



FOCuS Algorithm
Adaptive Maxima Checking

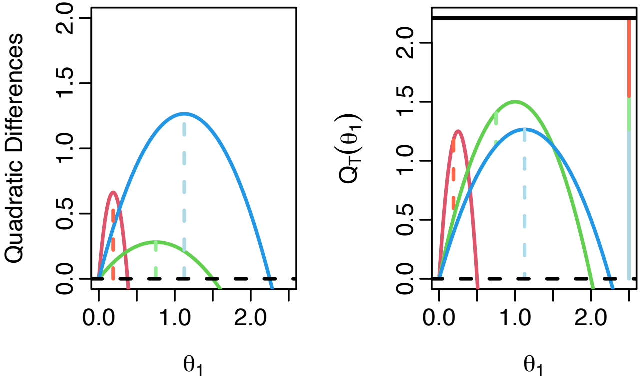
The main computational cost of FOCuS is calculating the maximum of $P_t(\mu)$, as this involves maximising each curve.

We can reduce this cost substantially by recycling calculations.

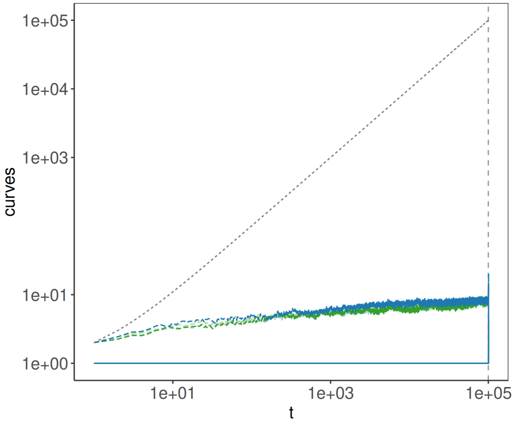
The key idea is that the difference between curves is unchanged as we add new curves. And we can use the maximum of these differences to bound the overall cost.

We start maximising curves from the most recent to the oldest. But, once our bound tells us that $\max P_t(\mu)$ is below our threshold we can stop.

FOCuS Algorithm
Adaptive Maxima Checking



FOCuS Algorithm
Adaptive Maxima Checking



FOCuS is an algorithm for online change detection that can efficiently calculate the Likelihood-ratio test statistic for all possible change locations.

- Essentially the same algorithm applies to any one-parameter exponential family model for the data.
- It can deal with both pre-change parameter known or unknown.
- With adaptive maxima checking it (empirically) has a constant per-iteration cost (that it is roughly equal to finding the maximum of one curve).
- The Poisson version of the algorithm is due to be used within the HERMES cube-satellite software for detecting gamma ray bursts.

Gamma-ray bursts detection with FOCuS-Poisson

Giuseppe Dilillo (INAF-IAPS)
Statistical Scalability for Data Streams,
London
2023-04-20

Gamma-ray bursts

What and why

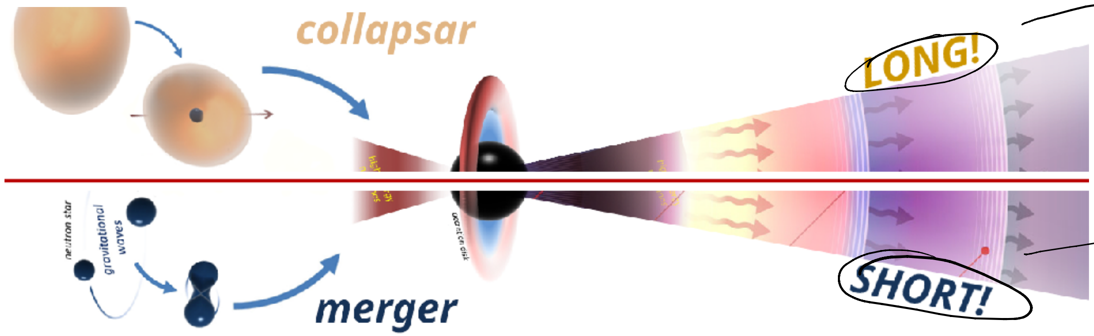
- Sudden bursts of energetic light.
- Sources are distant, new-born black holes.
- Duration between milliseconds and hours.
- Most powerful class of cosmic explosions.

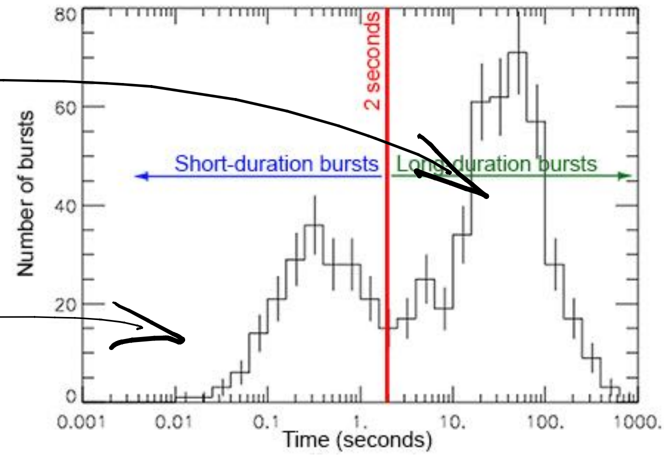
Support The Guardian
Available for everyone, funded by readers
Contribute → Subscribe →
My account
The Guardian

Science

New frontier for science as astronomers witness neutron stars colliding

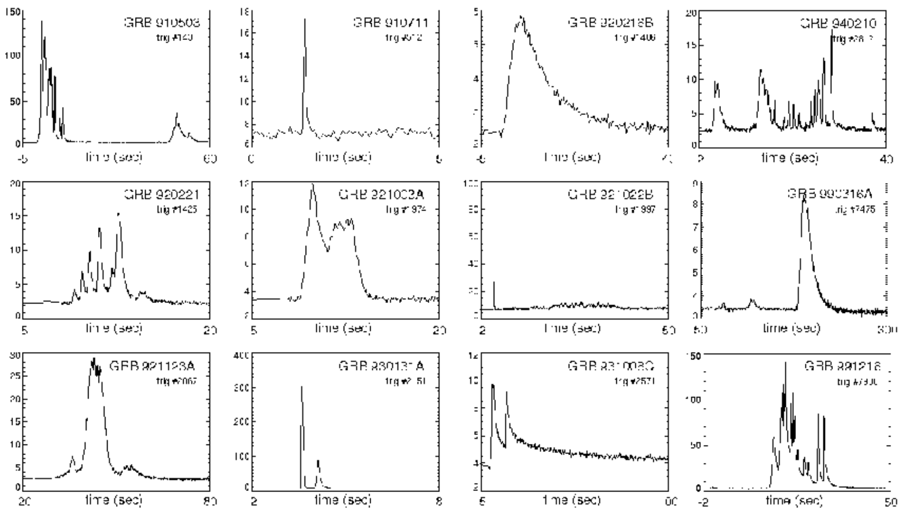
Extraordinary event has been 'seen' for the first time, in both gravitational waves and light - ending decades-old debate about where gold comes from



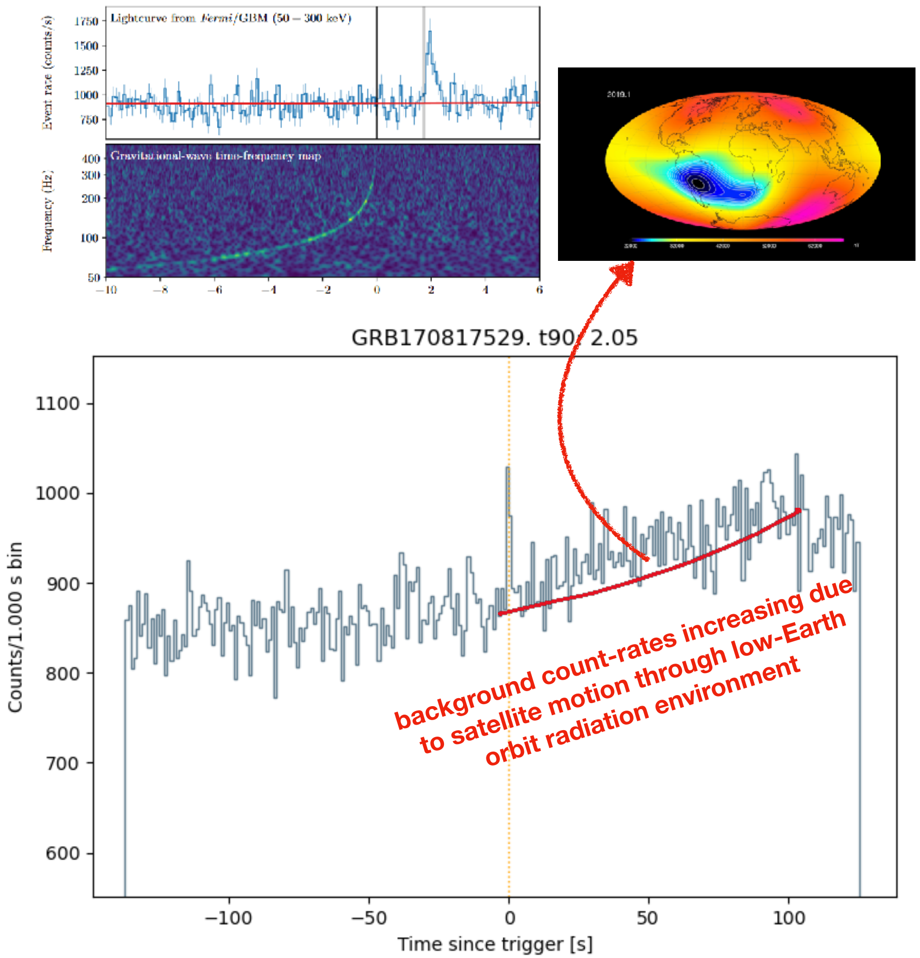


How GRBs look
and how to detect them

- From space, with count detectors.
- No two GRBs are the same.



Histograms of photons counts for extremely bright GRBs detected by BATSE aboard NASA Compton. There is a whole zoo of shapes and durations.
All of them do appear as a temporary change in rate over a poissonian background.



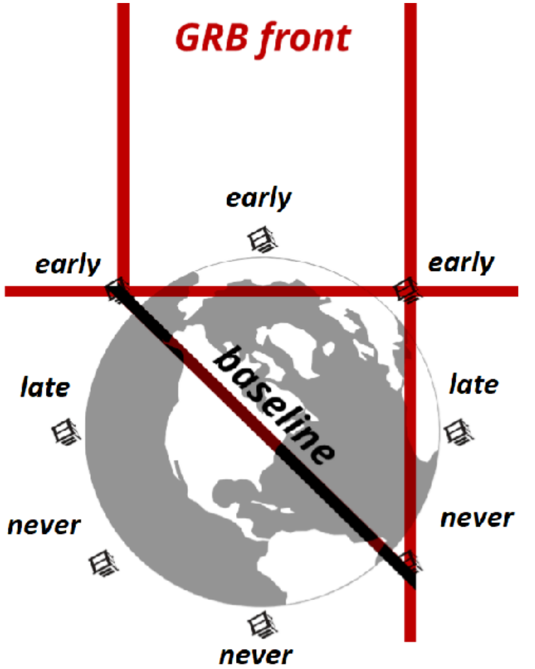
HERMES

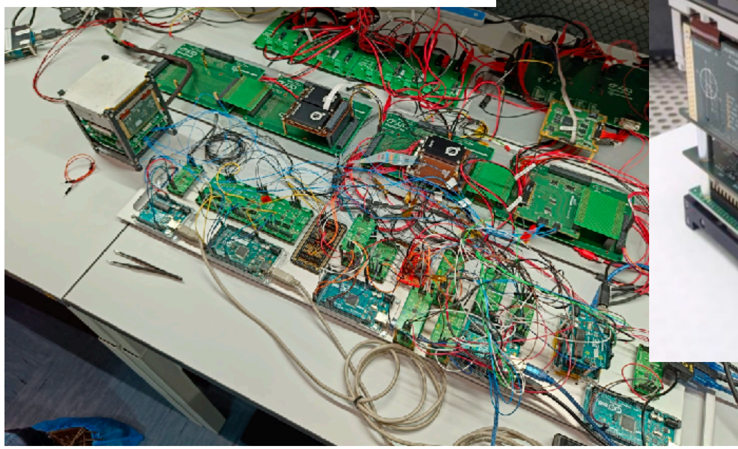
High-energy rapid modular ensemble of satellites

I work on this mission.

- 6+1 CubeSat constellation.
- GRB all-sky monitor.
- Source localization comparing photons arrival times between different spacecrafts.
- HORIZON2020/ASI funding.

https://www.hermes-sp.eu/





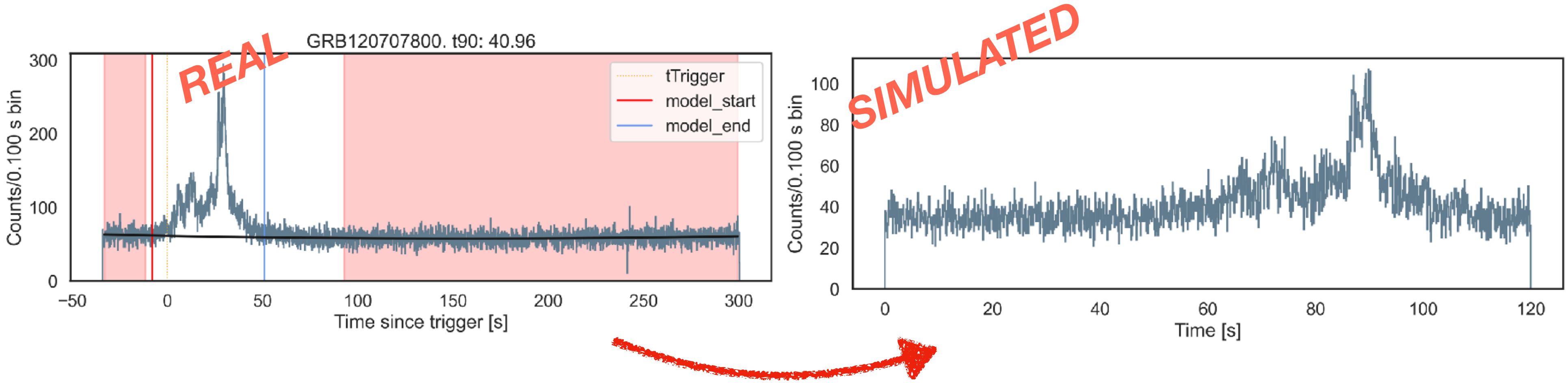
FOCuS and HERMES

1. Online burst search

- Embedded implementation for multiple detectors and energy bands.
- Background assessment and forecast through SMA or exponential smoothing.
- Tuned backtesting over archival Fermi-GBM dataset.
- Tested over a large library of synthetic GRBs.

Online burst search

Detection Performances



- We made a software to generate custom, synthetic GRB lightcurves modelled after real observations (Synthburst).
- Using Synthburst we generated a large library of synthetic lightcurves.
- We ran different algorithms and compared performances.

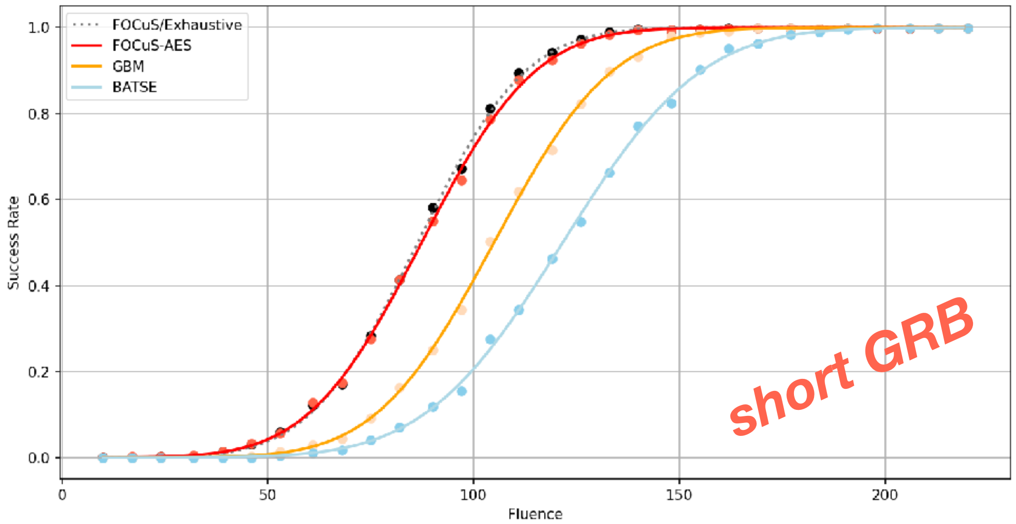
Online burst search

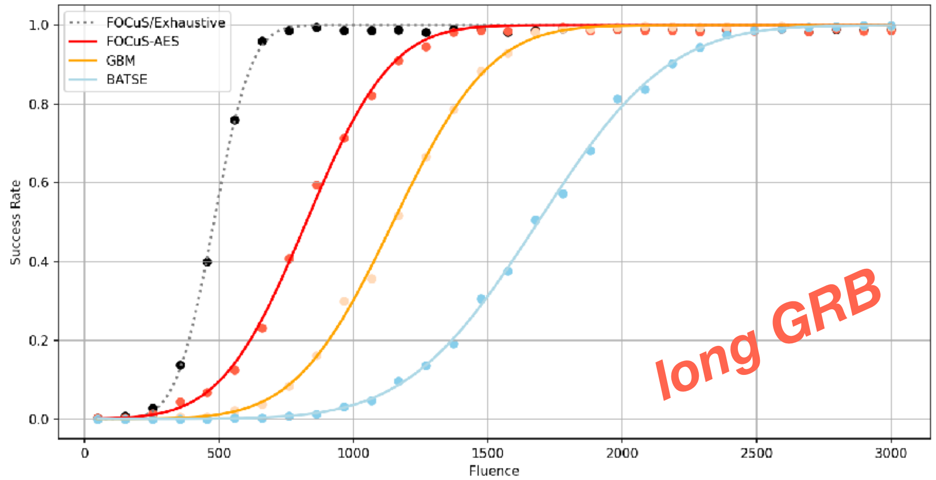
Detection Performances

- Finally, we compared detection performances from different algorithms.

Results:

- **Ideal detection performances** for approximated FOCuS-Poisson implementation over short bursts.
- Best performances over long bursts yet **limited by automatic background assessment.**
- **Half the computational cost** of the best benchmark we could come up with.





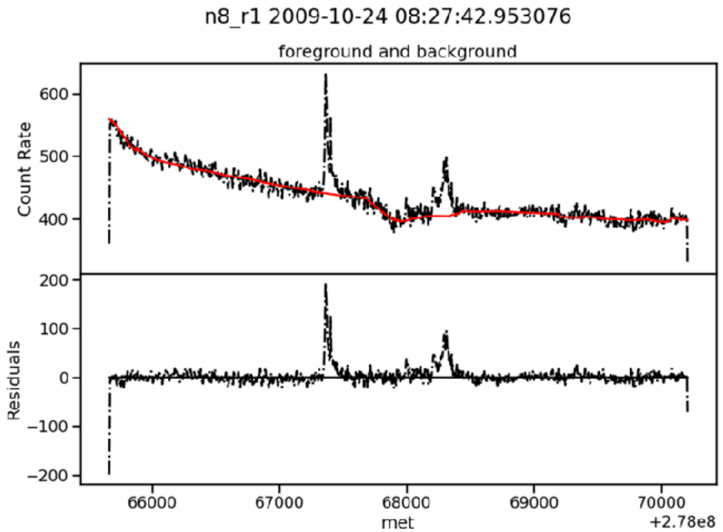
FOCuS and HERMES

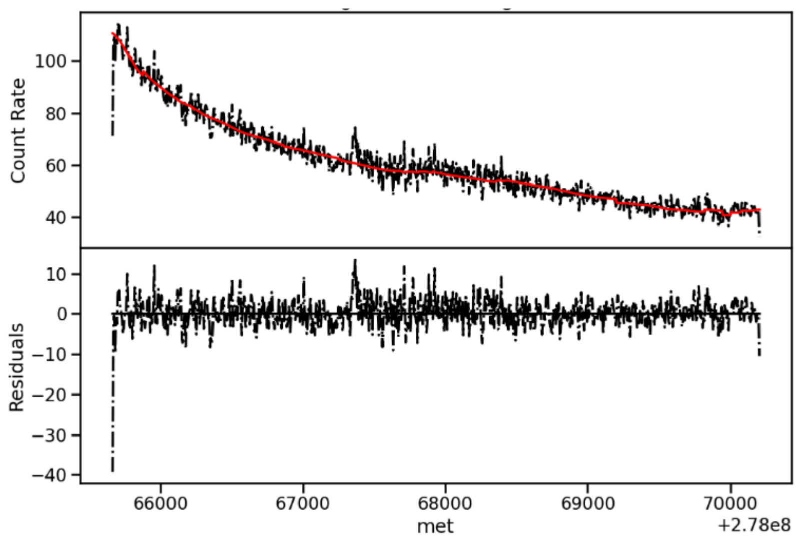
2. Offline burst search

- ML background model trained over archival data from GRB experiment (Riccardo Crupi).
- Used FOCuS to detect transients.
- We found multiple transients which were previously not known!

ArXiv: https://arxiv.org/pdf/2303.15936

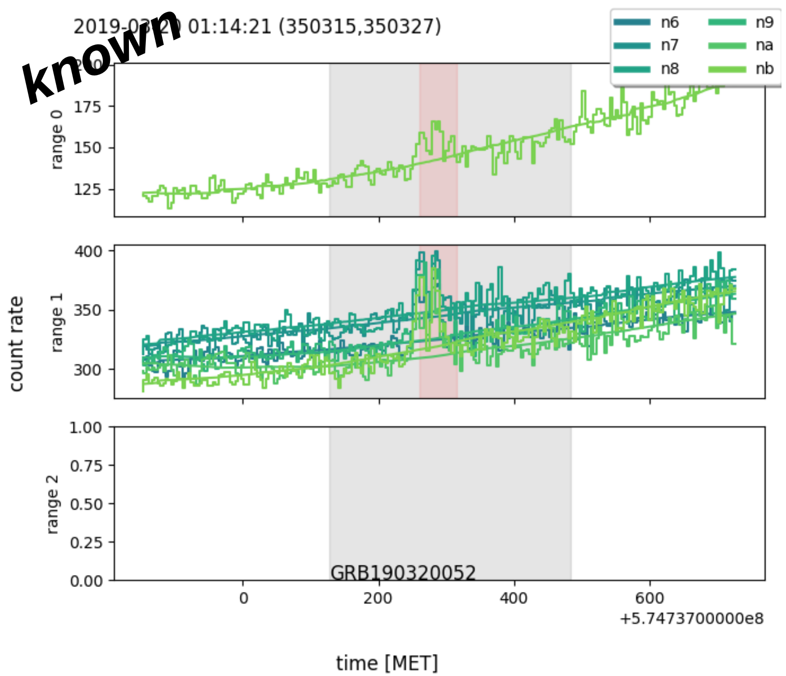
github: https://github.com/rcrupi/deepgrb

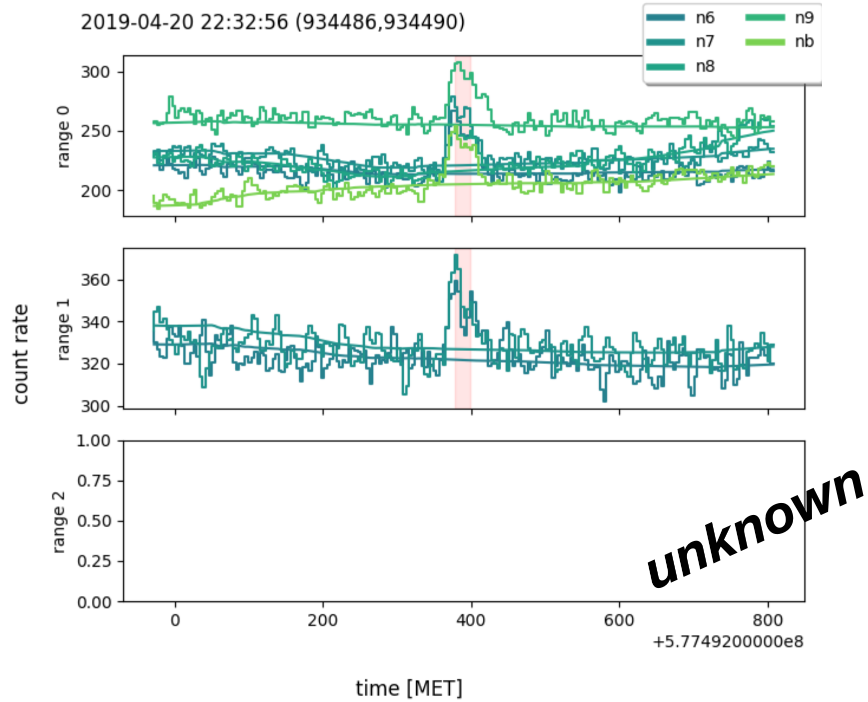




FOCuS and HERMES

2. Offline burst search





#	ID	Trigger time	T (s)	Detectors triggered	Catalog trigger name	S r0	S r1	S r2	CE
1	2010_1	2010-11-02 14:16:36	8.19	n2	UNKNOWN: GRB/GF	3.99	3.38	0	P
6	2010_6	2010-11-11 13:04:23	32.77	n2	UNKNOWN: GRB	4.29	5.81	0	P
7	2010_7*	2010-11-11 18:58:17	16.38	n2 n4 n5	UNKNOWN: SF	> 10	4.57	0	R
10	2010_10	2010-11-12 23:46:52	19.73	n1	UNKNOWN: GRB/GF	0	9.18	5.31	P
14	2010_14	2010-11-22 13:23:34	8.19	n5 nb	UNKNOWN: UNC(LP)/GRB	0	7.45	5.41	R
33	2010_34	2010-12-28 18:22:27	106.50	na	UNKNOWN: TGF	0	8.08	0	P

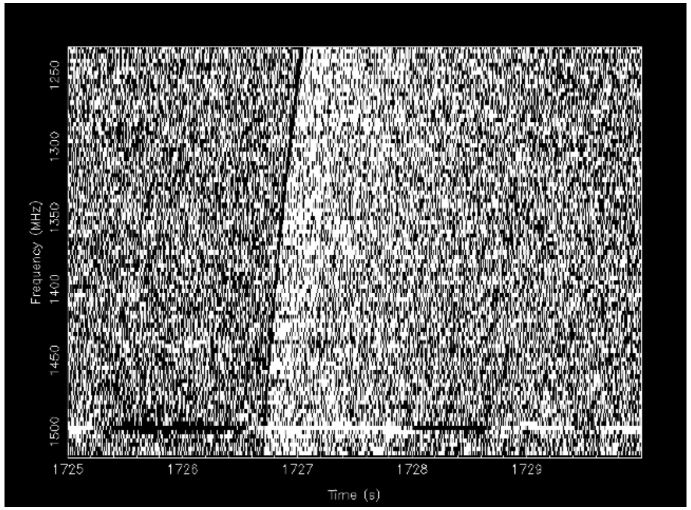
...

We made a small catalog of unknown events.

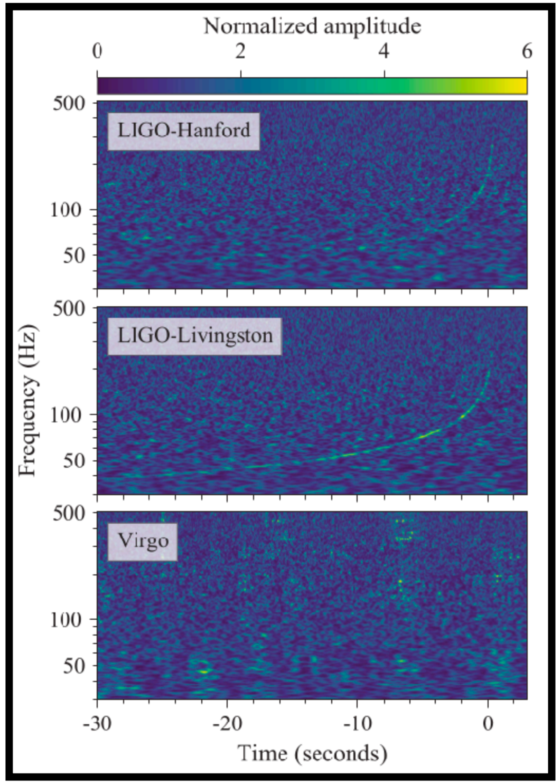
Beyond GRBs

Search for anomalies in streaming data is ubiquitous in astrophysics!

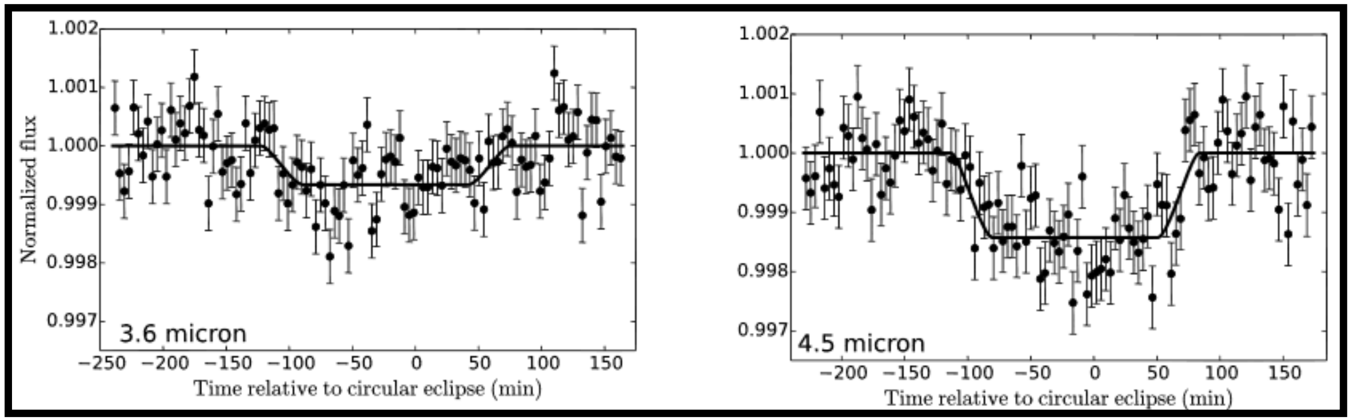
- Detect and characterize planetary and stellar occultations.
- Multi-dimensional search over frequency and energy domains.
- Images source detection.



First detected radio burst, Lorimet et al., 2013



GW170817, Abbott et al., 2017



Occultation from hot Jupyter exoplanet HAT-P-13b. From Buhler et al., 2013.

Thank you!

References

Crupi, Dilillo, Bissaldi, Fiore and Bari (2023) *Searching for long faint astronomical high energy transients: A data driven approach* arXiv:2303.15936

Page (1954) *Continuous inspection schemes* Biometrika

Romano, Eckley, Fearnhead and Rigaill (2023) *Fast Online Changepoint Detection via Functional Pruning CUSUM statistics.* Journal of Machine Learning Research

Ward, Dilillo, Eckley and Fearnhead (2022) *Poisson-FOCuS: A fast and efficient algorithm for detecting gamma ray bursts by cube-satellites.* arXiv:2208.01494

Ward, Romano, Eckley and Fearnhead (2023) *A Constant-per-Iteration Likelihood Ratio Test for Online Changepoint Detection for Exponential Family Models* arXiv:2302.04743

`R` code is available from `https://github.com/gtromano/FOCuS`; code for GRBs from `https://github.com/rcrupi/deepgrb`

FOCuS Algorithm
CPU cost

Why do we keep changes corresponding to the points on the convex hull?

There are two key properties

(1) If you join two points on the random walk – the slope of the line is the mean of the observations between those time-points.

(2) If you compare the likelihood for a change at s and one at t with $s < t$, then the earlier change is better for post-change means μ that are less than twice the mean of $y_{s+1:t}$; i.e.

$$\theta := \frac{\mu}{2} \leq \bar{y}_{s+1:t}$$

FOCuS Algorithm CPU cost

We can then work out what values of $\theta = \mu/2$ a change at τ is better than later changes:

$$\theta \leq \min_{t>\tau} \bar{y}_{\tau+1:t}.$$

And also better than earlier changes:

$$\theta \geq \max_{t<\tau} \bar{y}_{t:\tau}.$$

For there to be some $\theta > 0$ where both hold we need

$$0 \leq \max_{t<\tau} \bar{y}_{t:\tau} < \min_{t>\tau} \bar{y}_{\tau+1:t}.$$

FOCuS Algorithm
CPU cost