A new Deep Learning model for Gamma-Ray Bursts' light curves generation

R. Falco, N. Parmiggiani, A. Bulgarelli, G. Panebianco, M. Lombardi, L. Castaldini, A. Di Piano, V. Fioretti, C. Pittori, M. Tavani

INAF ISTITUTO NAZIONALE DI ASTROFISICA

OAS BOLOGNA

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION CATANIA, 8-12 JULY, 2024

Gamma-Ray Bursts

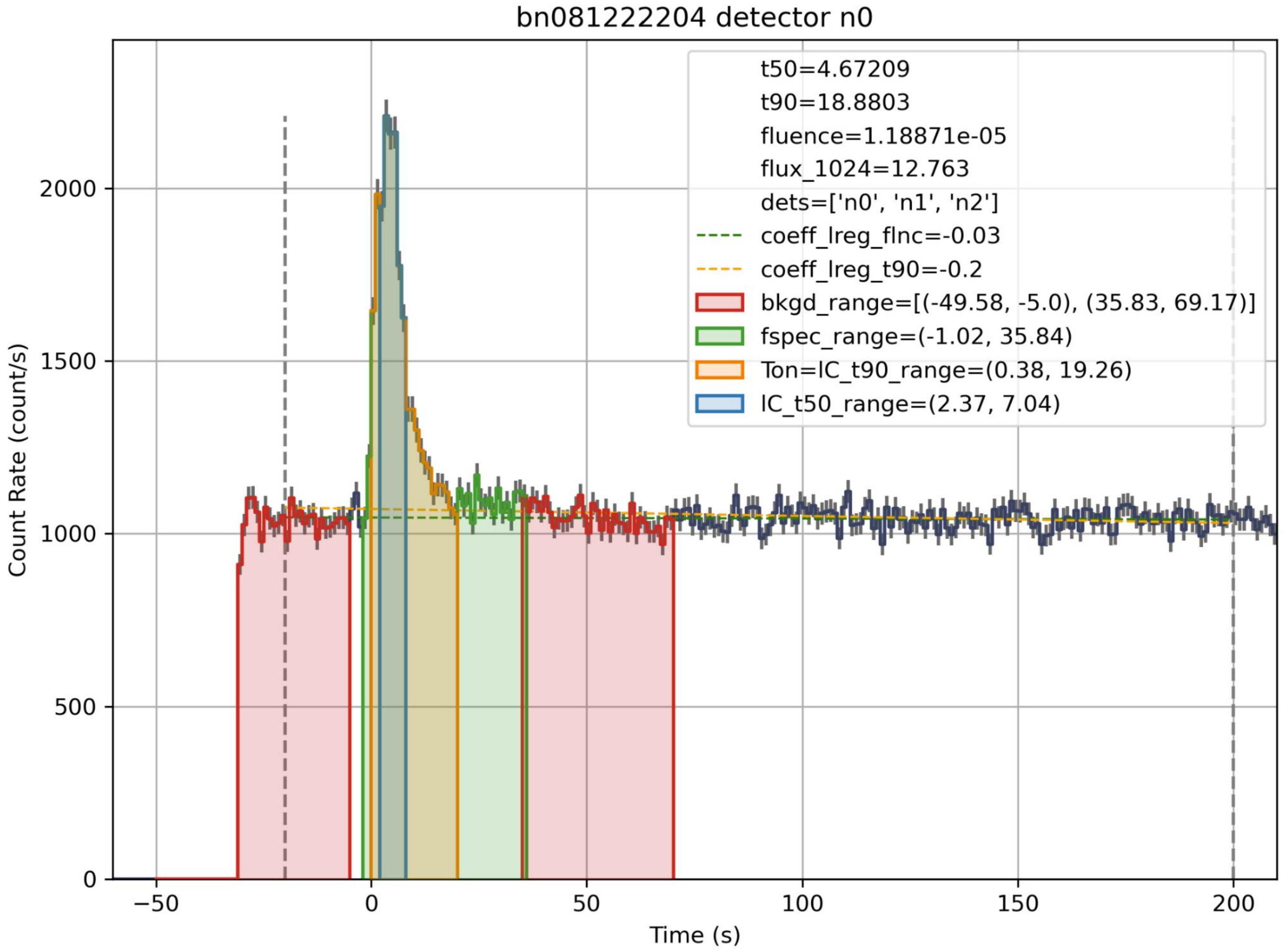
- GRBs are events that can generate more energy in 10 seconds than the Sun can in its entire lifetime
- They are **transient events** that can last from milliseconds to several hours
- Can be produced by **various sources**
  - *Neutron stars mergers*
  - *Collapse of massive stars*

Several space missions (like **AGILE** and **Fermi**) and ground-based instruments work to discover new GRB every day

image: http://go.nasa.gov/17R2cf3

Light curves

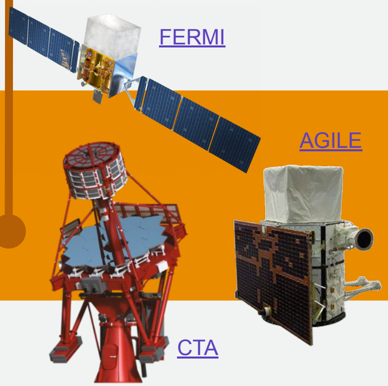
- Data are considered as **time series**
- Light curves are plot representing aggregated counts in bins with a **certain time interval**
- It is then used to **analyze the temporal trend** of a gamma-ray source



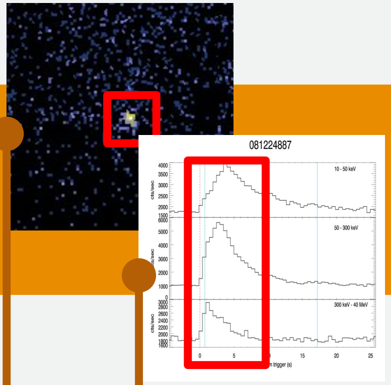
GRBs detection and analysis

Source activity (GRB)

Terrestrial and space mission



Train and test GRB detection methods with **standard** and **machine learning tools**



Skymap detection

Light curve detection

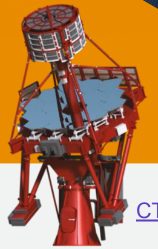
High energy astrophysics analysis

GRBs detection and analysis

Source activity
(Burst of γ-ray)

Terrestrial and
space mission

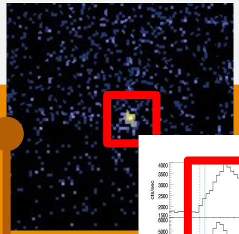
FERMI



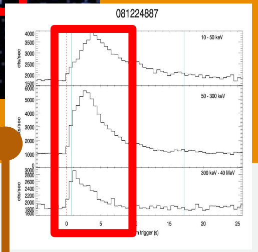
CTA

AGILE

Train and test GRB detection methods with **standard** and **machine learning tools**



Skymap
detection



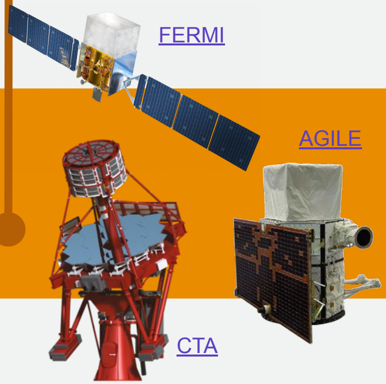
Light curve
detection

High energy
astrophysics analysis

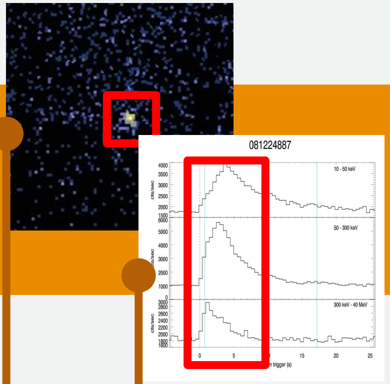
GRBs detection and analysis

Source activity
(Burst of γ-ray)

Terrestrial and
space mission

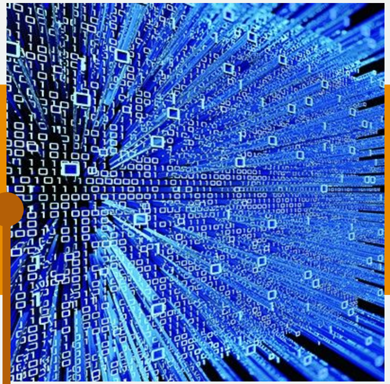


Train and test GRB detection methods with **standard** and **machine learning tools**



Skymap
detection

Light curve
detection



NEED OF LARGE AMOUNT OF DATA

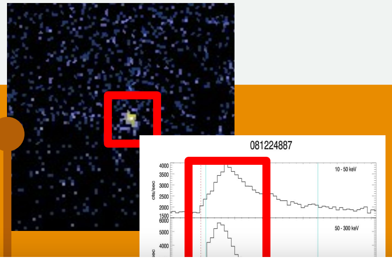
High energy
astrophysics analysis

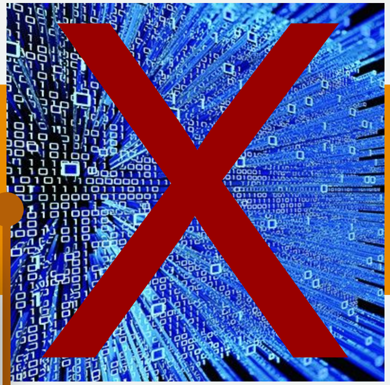
GRBs detection and analysis

Terrestrial and space mission

Train and test GRB detection methods with **standard** and **machine learning tools**

High energy astrophysics analysis





Unfortunately, GRBs are **rare phenomena** so they are difficult to obtain.

Source activity (Burst of γ-ray)

Light curve detection

Skymap detection

NEED OF LARGE AMOUNT OF DATA

GRBs detection and analysis

Cornell University

We gratefully acknowledge

arXiv > astro-ph > arXiv:2002.11460

Search...

Help | Adv

Astrophysics > High Energy Astrophysical Phenomena

[Submitted on 26 Feb 2020 (v1), last revised 14 Apr 2020 (this version, v2)]

The Fourth Fermi-GBM Gamma-Ray Burst Catalog: A Decade of Data

A. von Kienlin, C. A. Meegan, W. S. Paciesas, P. N. Bhat, E. Bissaldi, M. S. Briggs, E. Burns, W. H. Cleveland, M. H. Gibby, M. M. Giles, A. Goldstein, R. Hamburg, C. M. Hui, D. Kocevski, B. Mailyan, C. Malacaria, S. Poolakkil, R. D. Preece, O. J. Roberts, P. Veres, C. A. Wilson-Hodge

We present the fourth in a series of catalogs of gamma-ray bursts (GRBs) observed with Fermi's Gamma-Ray Burst Monitor (Fermi-GBM). It extends the six year catalog by four more years, now covering the ten year time period from trigger enabling on 2008 July 12 to 2018 July 11. During this time period GBM triggered almost twice a day on transient events of which we identifyied 2356 as cosmic GRBs. Additional trigger events were due to solar are events, magnetar burst activities, and terrestrial gamma-ray flashes. The intention of the GBM GRB catalog series is to provide updated information to the community on the most important observables of the GBM-detected GRBs. For each GRB the location and main characteristics of the prompt emission, the duration, peak flux, and fluence are derived. The latter two quantities are calculated for the 50-300 keV energy band, where the maximum energy release of GRBs in the instrument reference system is observed and also for a broader energy band from 10-1000 keV, exploiting the full energy range of GBM's low-energy detectors. Furthermore, information is given on the settings of the triggering criteria and exceptional operational conditions during years 7 to 10 in the mission. This fourth catalog is an official product of the Fermi-GBM science team, and the data files containing the complete results are available from the High-Energy Astrophysics Science Archive Research Center (HEASARC).

Comments:	273 pages, 10 figures, 8 tables. This is a 10 year catalog update of arXiv:1603.07612
Subjects:	**High Energy Astrophysical Phenomena (astro-ph.HE)**
Cite as:	arXiv:2002.11460 **[astro-ph.HE]**
	(or arXiv:2002.11460v2 **[astro-ph.HE]** for this version)

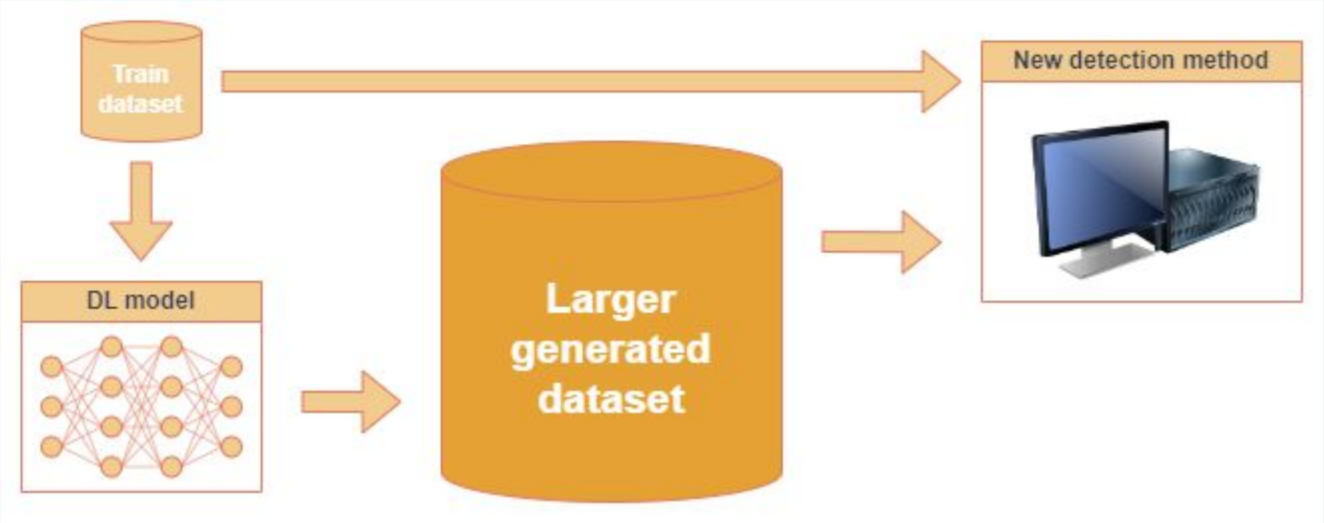
ref: The Fourth Fermi-GBM Gamma-Ray Burst Catalog: A Decade of Data

A DL model to simulate GRBs

We developed a **Deep Learning** model to generate GRBs’ LCs

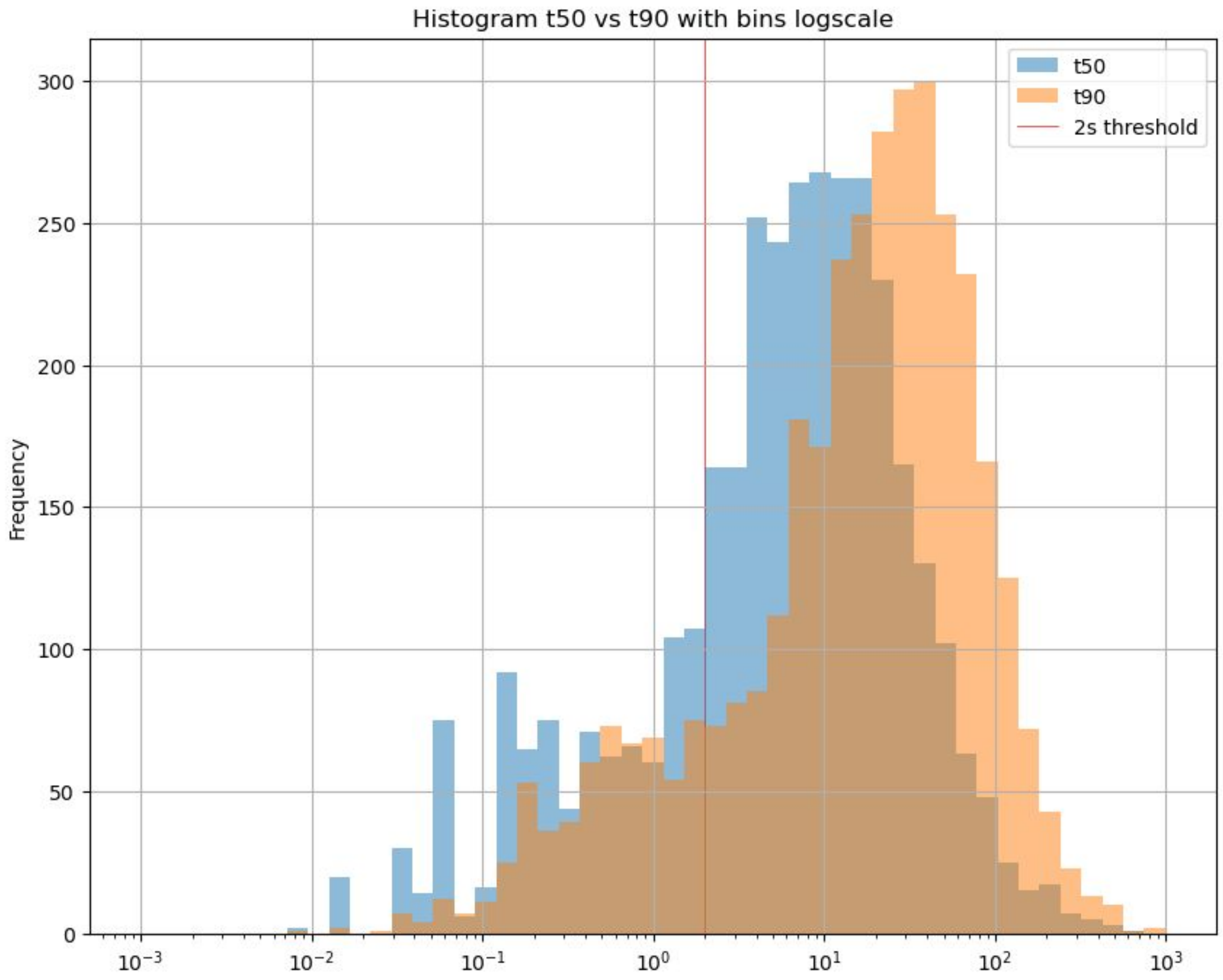
Obtain **larger dataset of LCs**

Train and evaluate new GRBs detection methods



Preprocessing

- LCs obtained from *4th Fermi-GBM GRB Catalog*
  - *3.608 GRBs, 11.869 LCs*
- LCs of the same GRB detected by multiple detectors are **independent time series**
  - *as data augmentation*
- We selected only **long GRBs** (t_{90} > 2s)
- We extracted LCs of **220 seconds** and **bins of 1 second**
- All the light curves time are in the range [-20, 200]
  - *respect to the trigger time t_0*



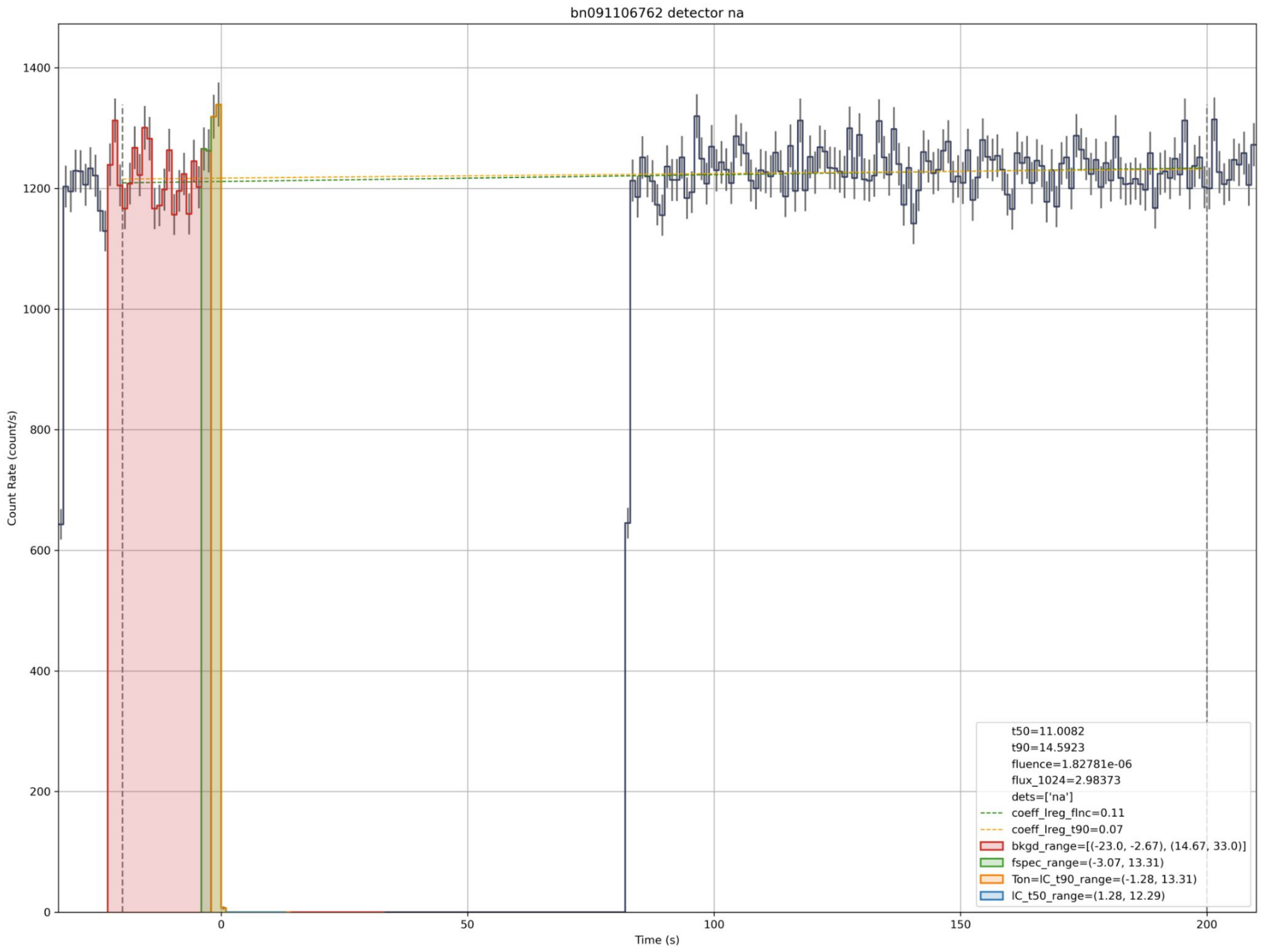
t90a nd t50 distribution, with double bell shape. Red line divide the classification of long and short GRBs.

Preprocessing

Filters to drop outliers:

- **F1**: filter only **long GRBs** (with t_{90}>2s)
- **F2**: filter only the GRBs for which **there exists at least one NaI detector**
- **F3**: filter only the LCs with **no missing data in between the t90 time interval**
- …

	base	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*
num LCs	11869	9896	9867	9822	9064	7099	5964
num GRBs	3608	3012	3005	2992	2761	2486	2233



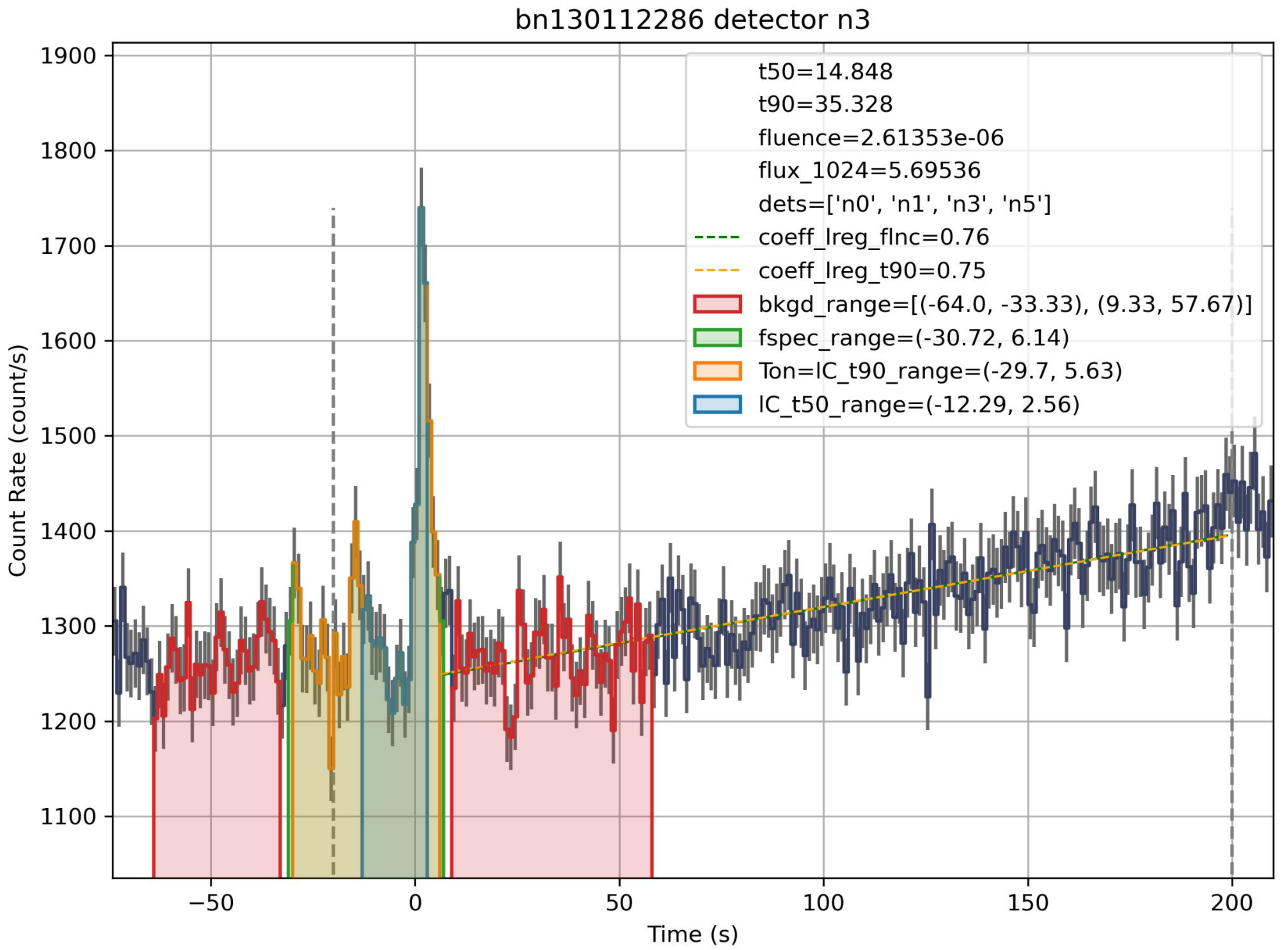
Drop LCs with **missing values in the burst flow range**

Preprocessing

Filters to drop outliers:

- **F1**: filter only **long GRBs** (with t_{90}>2s)
- **F2**: filter only the GRBs for which **there exists at least one NaI detector**
- **F3**: filter only the LCs with **no missing data in between the t90 time interval**
- **F4**: filter LCs with an estimated angular coefficient of LR *-0.4 ≤ coef ≤ 0.4*
- …

	base	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*
num LCs	11869	9896	9867	9822	9064	7099	5964
num GRBs	3608	3012	3005	2992	2761	2486	2233



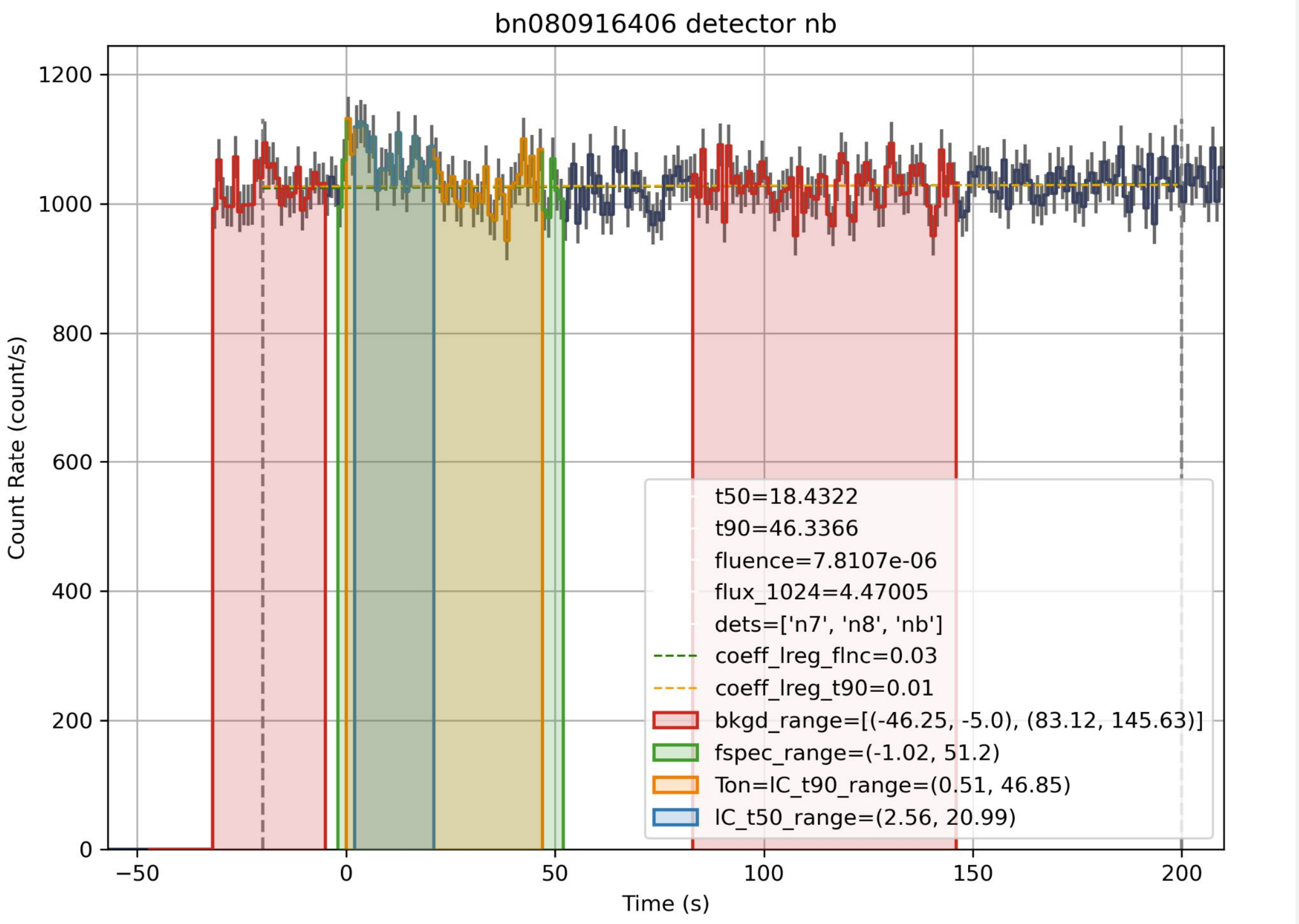
Drop LCs with **strange trends in background**, using a Linear Regressor (LR) fit on background regions with LR's **angular coefficient** *coef* s.t. *coef* < *-0.4* ∧ *coef* > *0.4*

Preprocessing

Filters to drop outliers:

- **F1**: filter only **long GRBs** (with t_{90}>2s)
- **F2**: filter only the GRBs for which **there exists at least one NaI detector**
- **F3**: filter only the LCs with **no missing data in between the t90 time interval**
- **F4**: filter LCs with an estimated angular coefficient of LR *-0.4 ≤ coef ≤ 0.4*
- **F5**: filter LCs, after computing Li&Ma GRB significance, with *σ > 3.0*
- …

	base	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*
num LCs	11869	9896	9867	9822	9064	7099	5964
num GRBs	3608	3012	3005	2992	2761	2486	2233



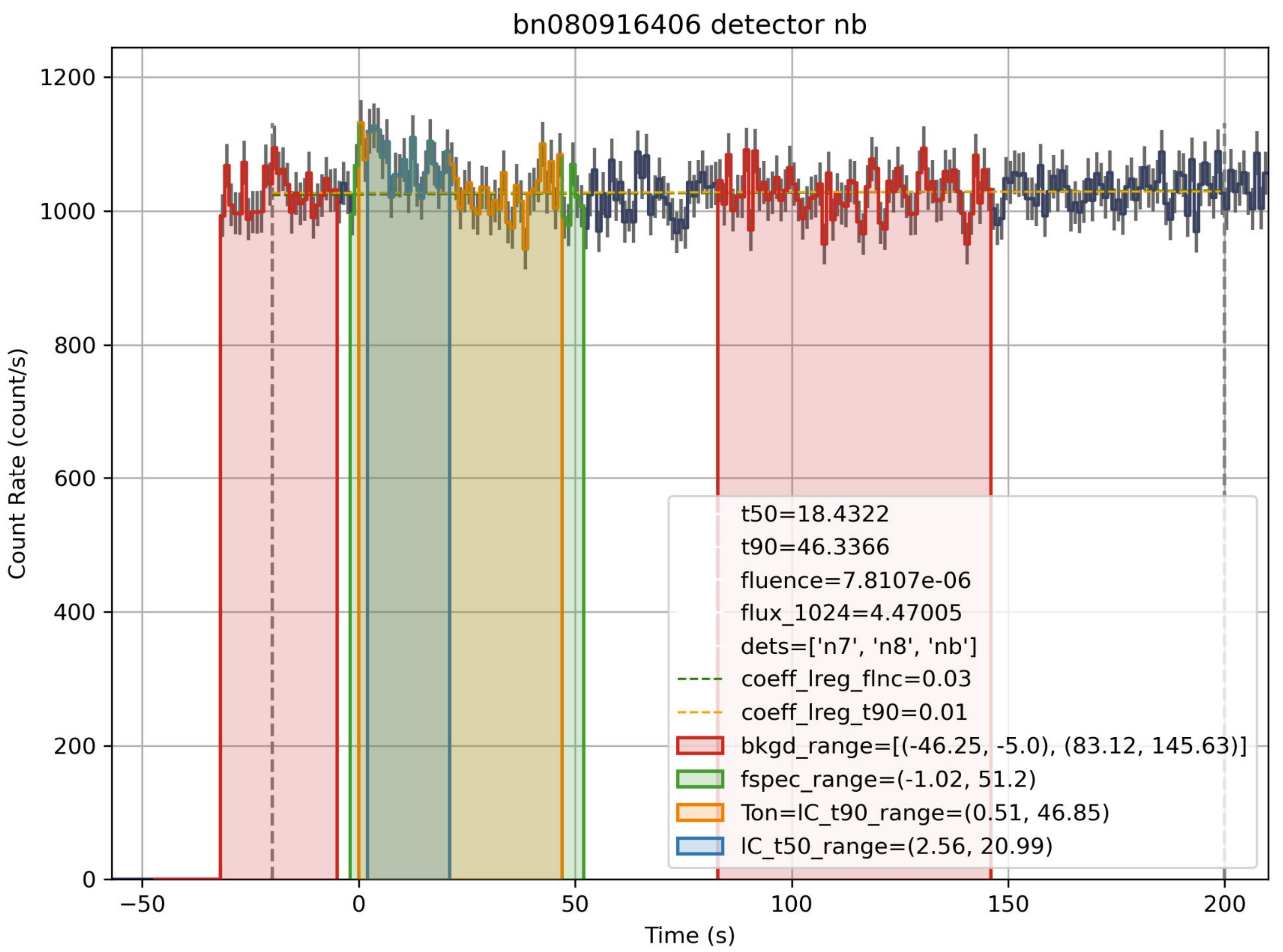
Drop LCs with **low peaks respect to background**, using Li&Ma GRB significance σ, by comparing photon count accumulation in the flow range with respect to photon count accumulation in the background ranges. Drop LCs with *σ ≤ 3.0*

Preprocessing

Filters to drop outliers:

- **F1**: filter only **long GRBs** (with t_{90}>2s)
- **F2**: filter only the GRBs for which **there exists at least one NaI detector**
- **F3**: filter only the LCs with **no missing data in between the t90 time interval**
- **F4**: filter LCs with an estimated angular coefficient of LR *-0.4 < coef < 0.4*
- **F5**: filter LCs, after computing Li&Ma GRB significance, with $\sigma > 3.0$
- **F6**: filter LCs from outliers with InterQuartile Range (IQR) method

	base	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*
num LCs	11869	9896	9867	9822	9064	7099	5964
num GRBs	3608	3012	3005	2992	2761	2486	2233



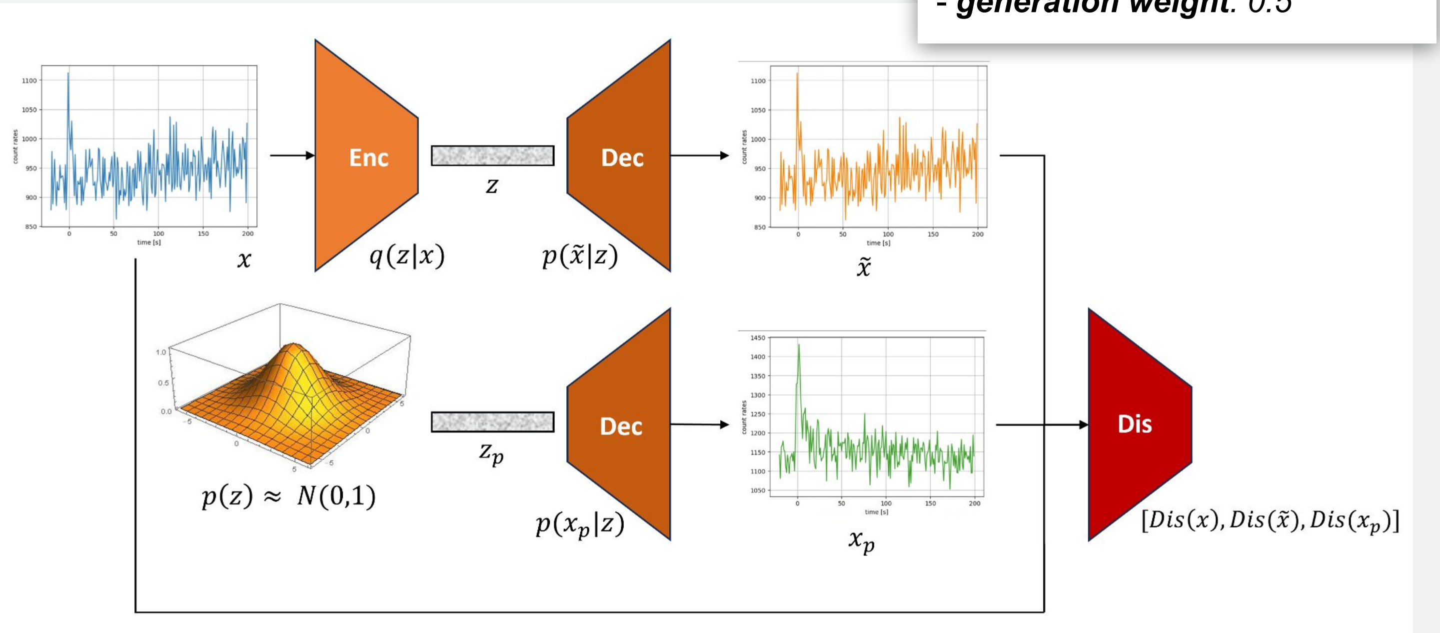
Drop LCs with **low peaks respect to background**, using Li&Ma GRB significance σ, by comparing photon count accumulation in the flow range with respect to photon count accumulation in the background ranges. Drop LCs with $\sigma \leq 3.0$

VAEGAN

- This DL model combines advantages of GANs and VAEs
  - *learns to encode and decode data*
  - *learns to compare dataset samples*
  - *this model produces high-quality and structured generative results while preserving latent space structure*
- 3 modules:
  - *Encoder*
  - *Decoder/Generator*
  - *Discriminator*
- The model is implemented using CNN layers to capture spatial local structures

Training recipe:

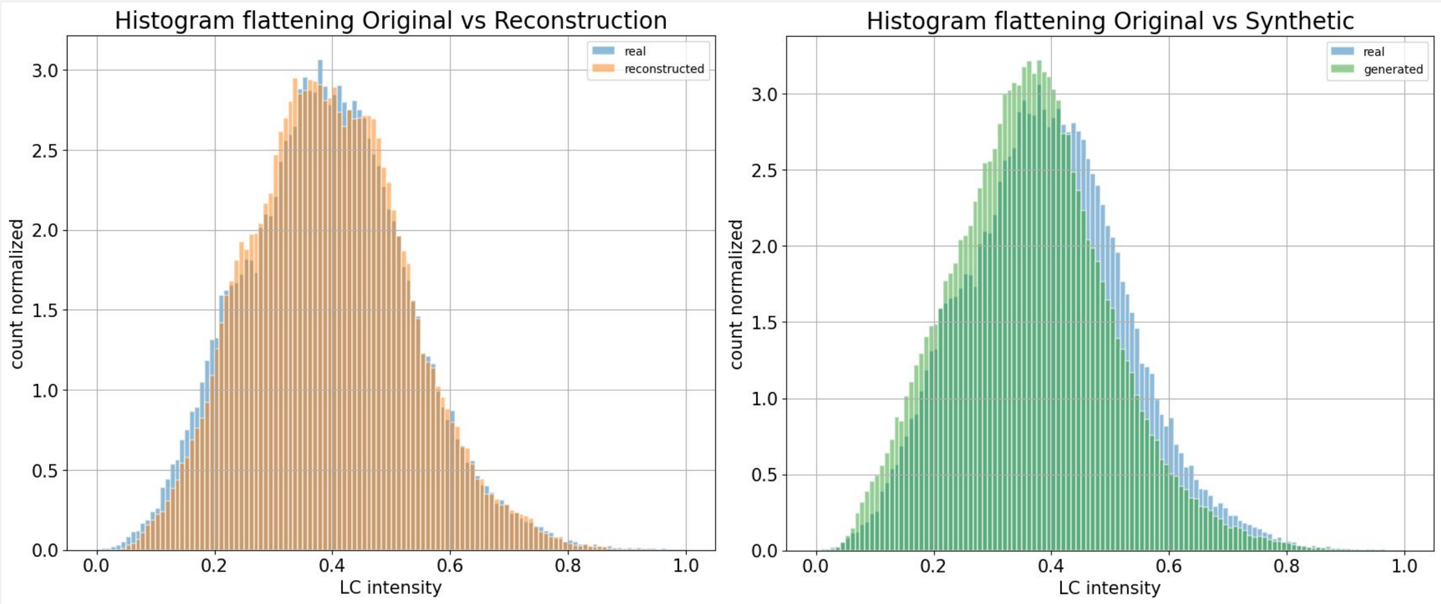
- *epochs: 1000*
- *batch size: 16*
- *Adam, lr: 5e-4*
- ***reconstruction weight**: 0.5*
- ***generation weight**: 0.5*

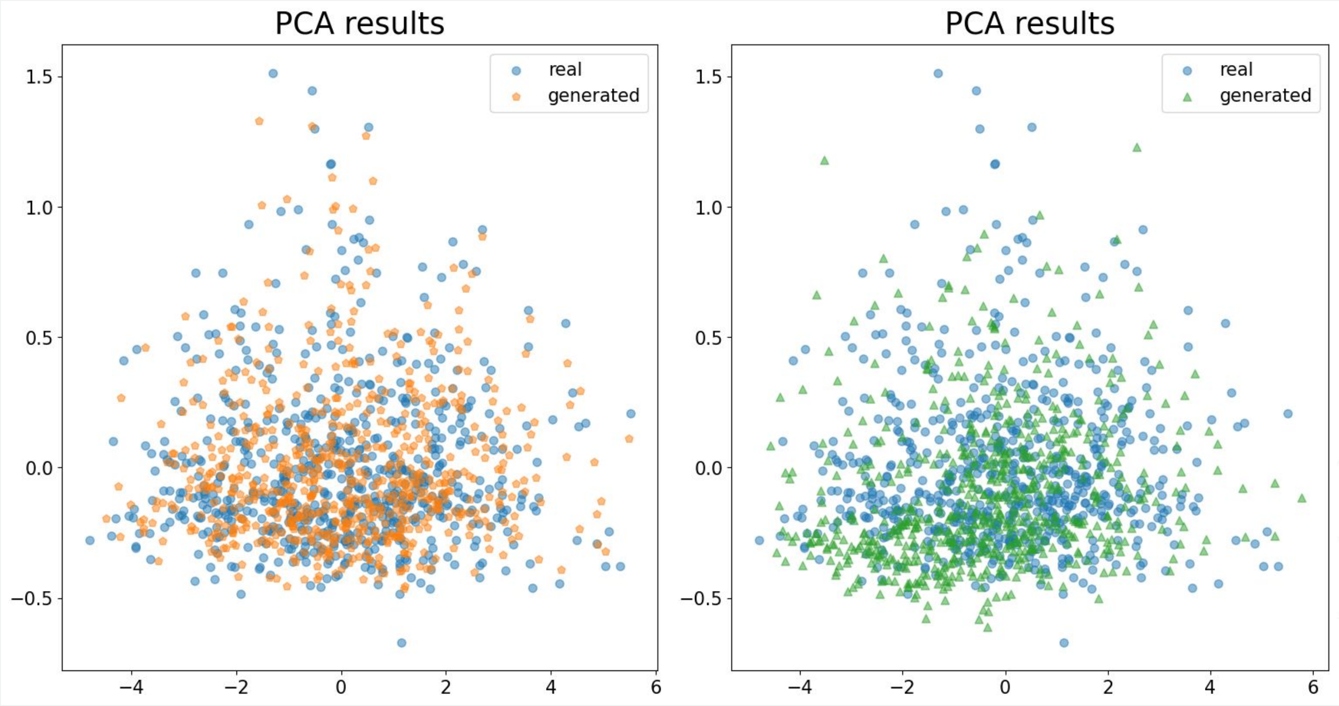


Larsen et al. in, 2016

Results

- ***Reconstruction***: count rates histograms **don't overlap perfectly** but it is still reasonable
  - L2 distance: 1.09
  - Wasserstein distance: 0.01
- ***Generation***:
  - *A **discrepancy** can be noticed.*
  - *The rates of the simulated LCs are shifted further to the left than those of the reals, so they tend to be **weaker***
  - *In **PCA** a **good dispersion** of the simulated LCs is evident*
  - *They are also **evenly distributed** with the real LCs (blue dots) **except in the bottom part of the plot***
    - L2 distance: 2.73
    - Wasserstein distance: 0.05

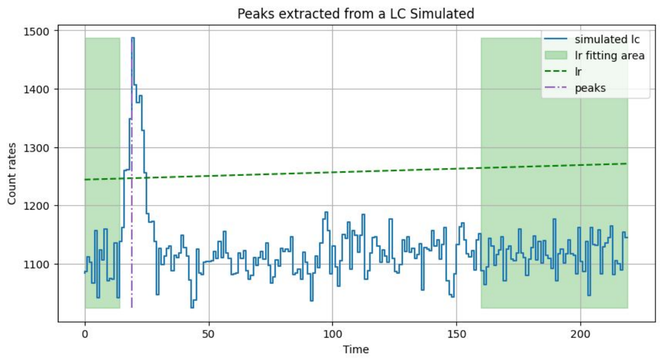




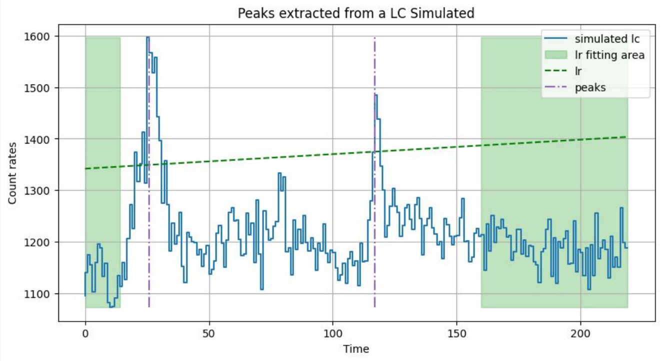
Number of peaks analysis

- Estimate the similarity between real samples and simulated data using dimensionality reduction techniques:
    - classify light curves by the **number of peaks** that compose them
    - ...

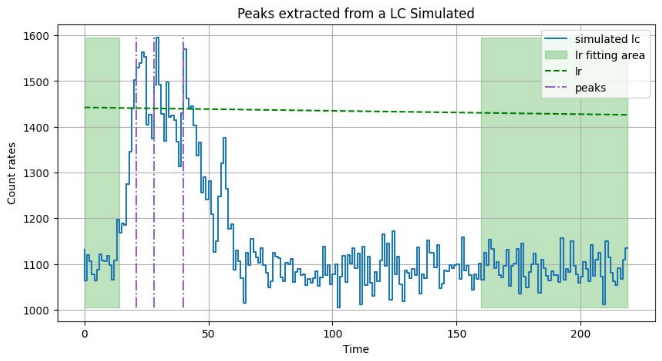
1 peak



2 peak

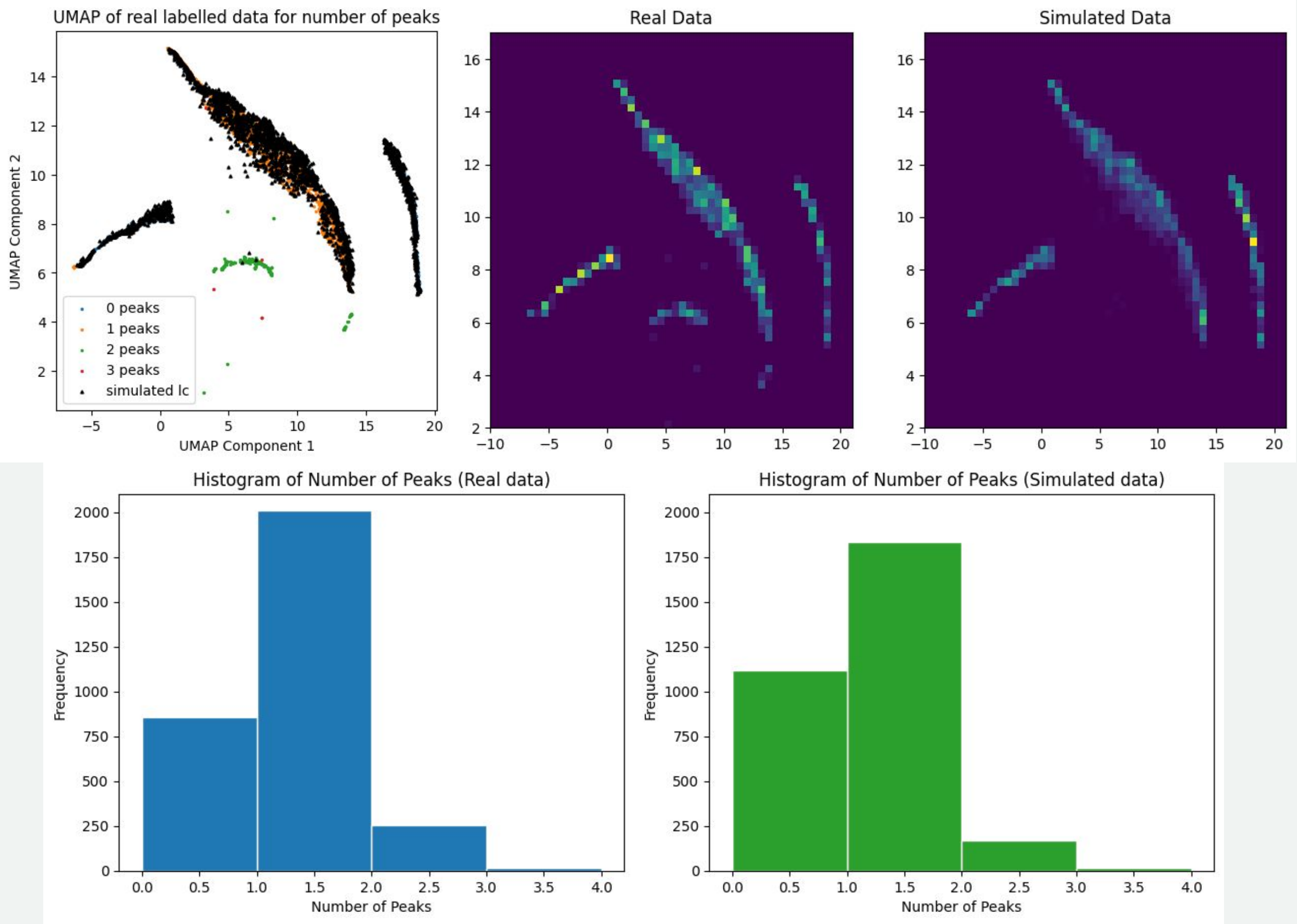


3 peak

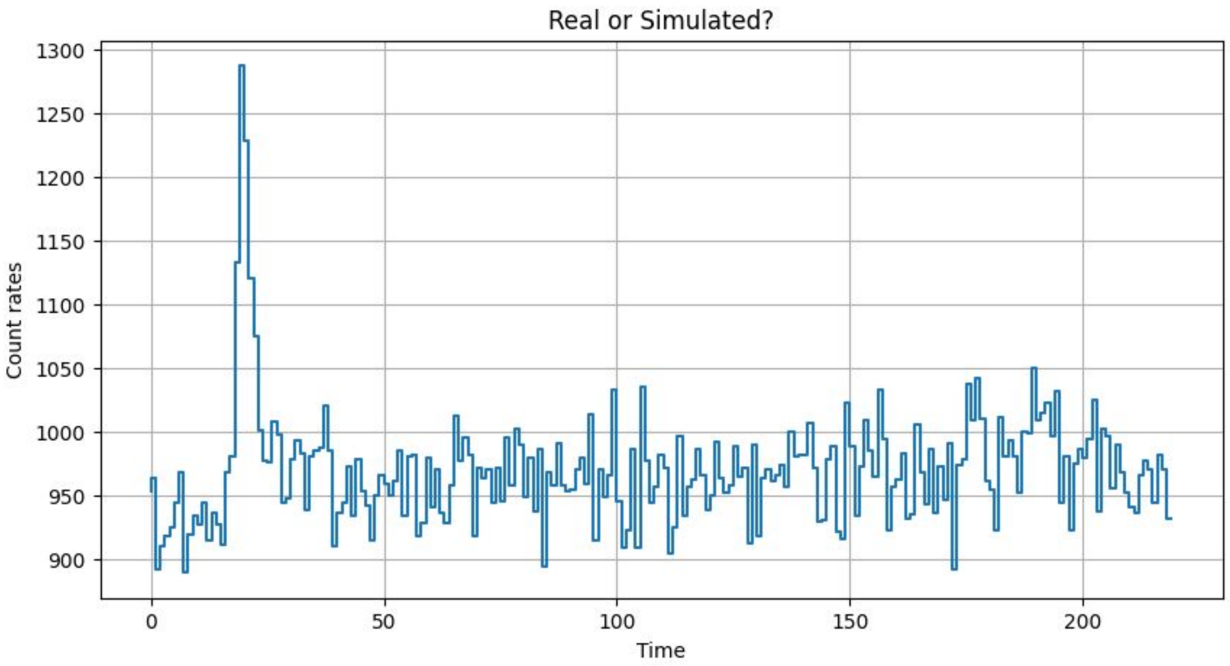


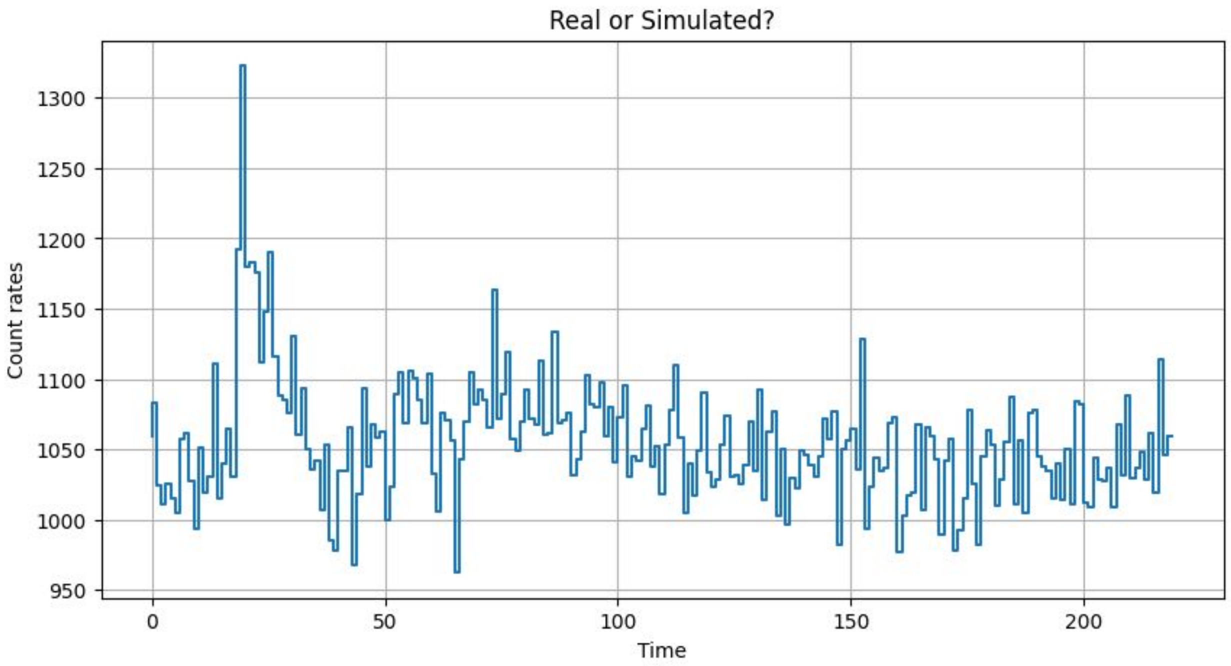
Number of peaks analysis

- Estimate the similarity between real samples and simulated data using dimensionality reduction techniques:
  - classify light curves by the **number of peaks** that compose them
  - **fit a dimensionality reduction technique on real data** to find clusters of data that have the same labels in common
  - plot the reduced simulated data and analyze the distances with the labeled real data



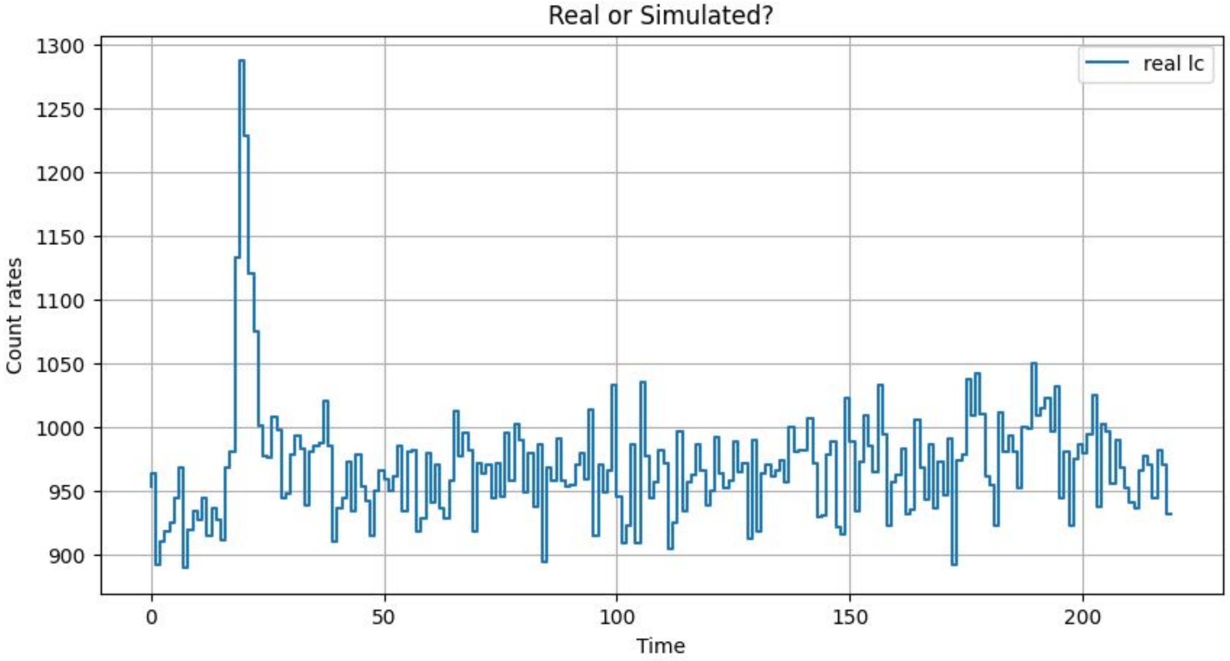
Simulated or real GRBs?



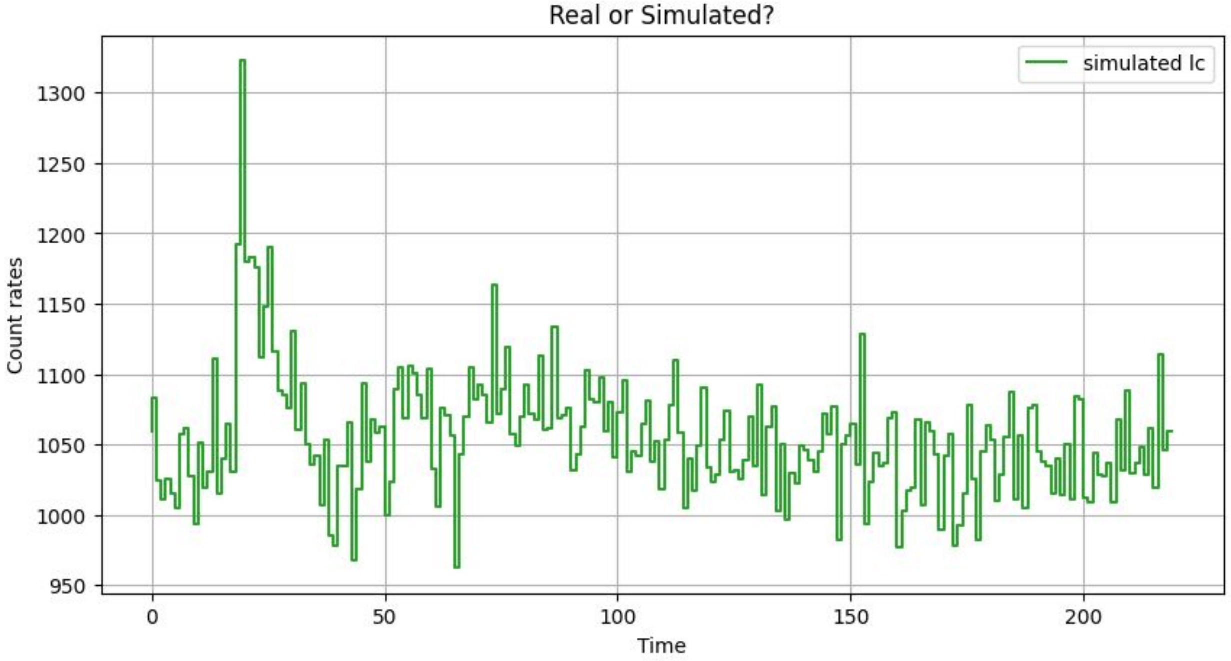


Simulated or real GRBs?

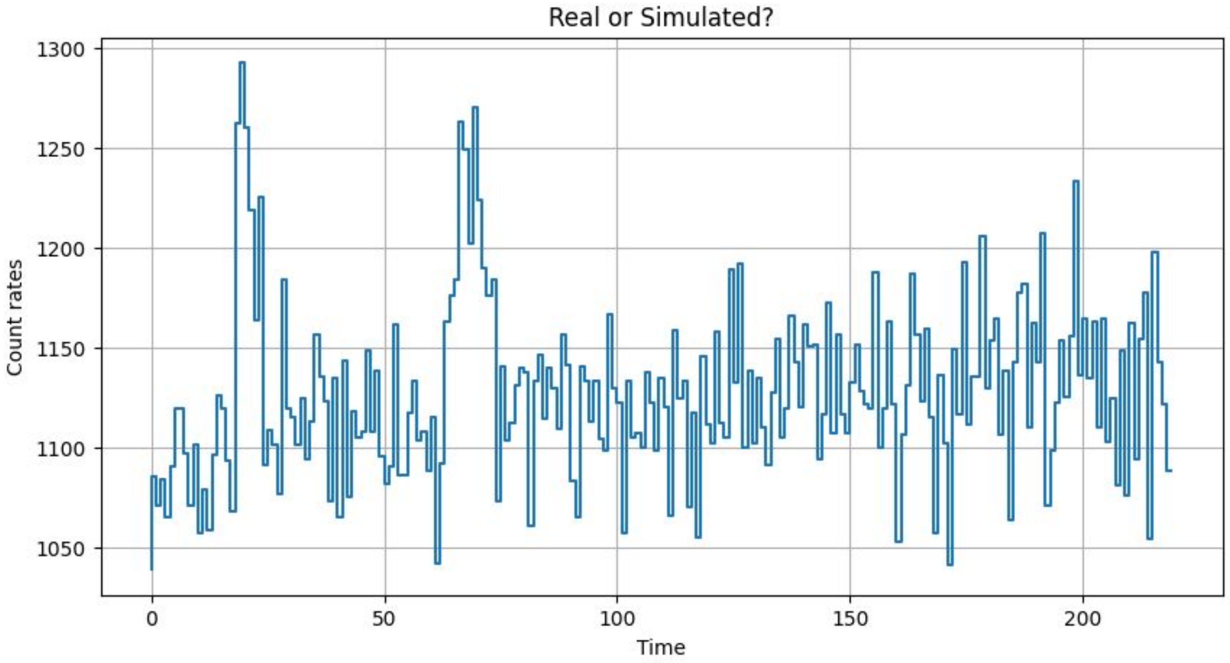
Real LC

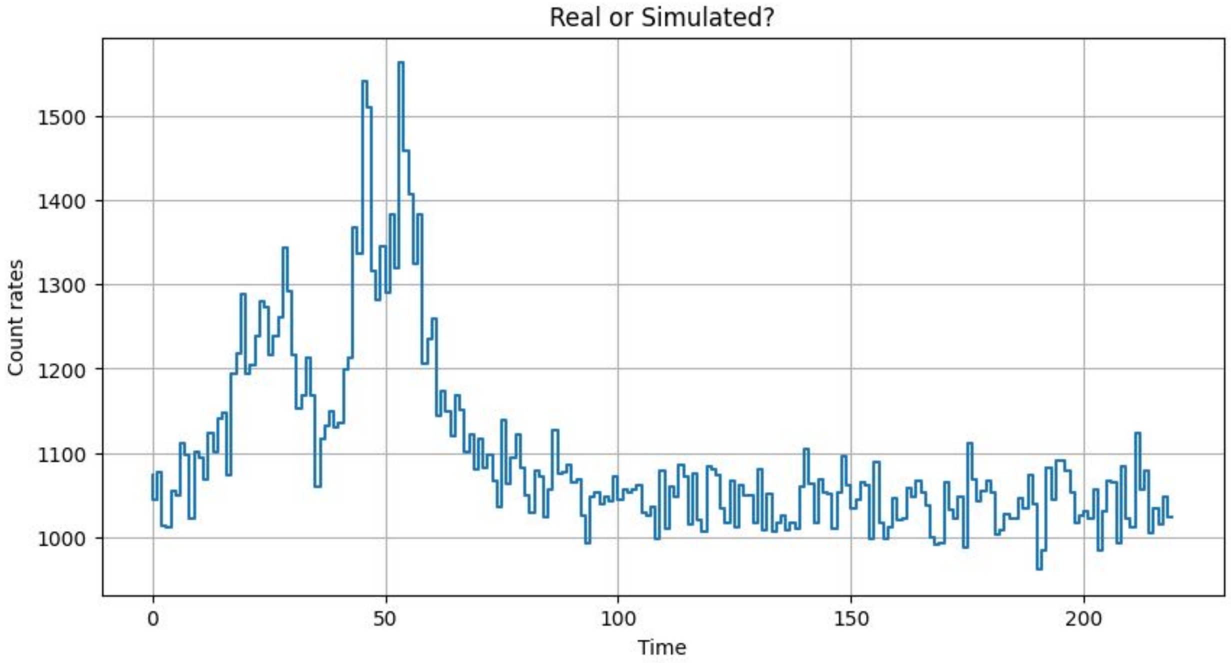


Simulated LC



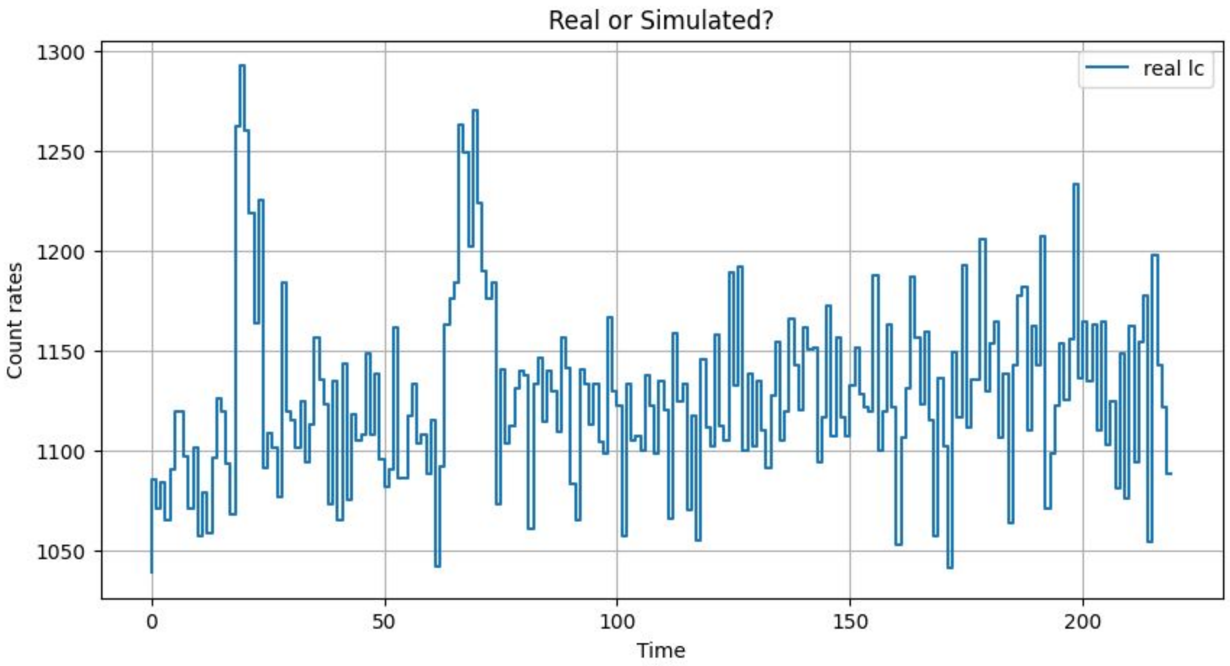
Simulated or real GRBs?



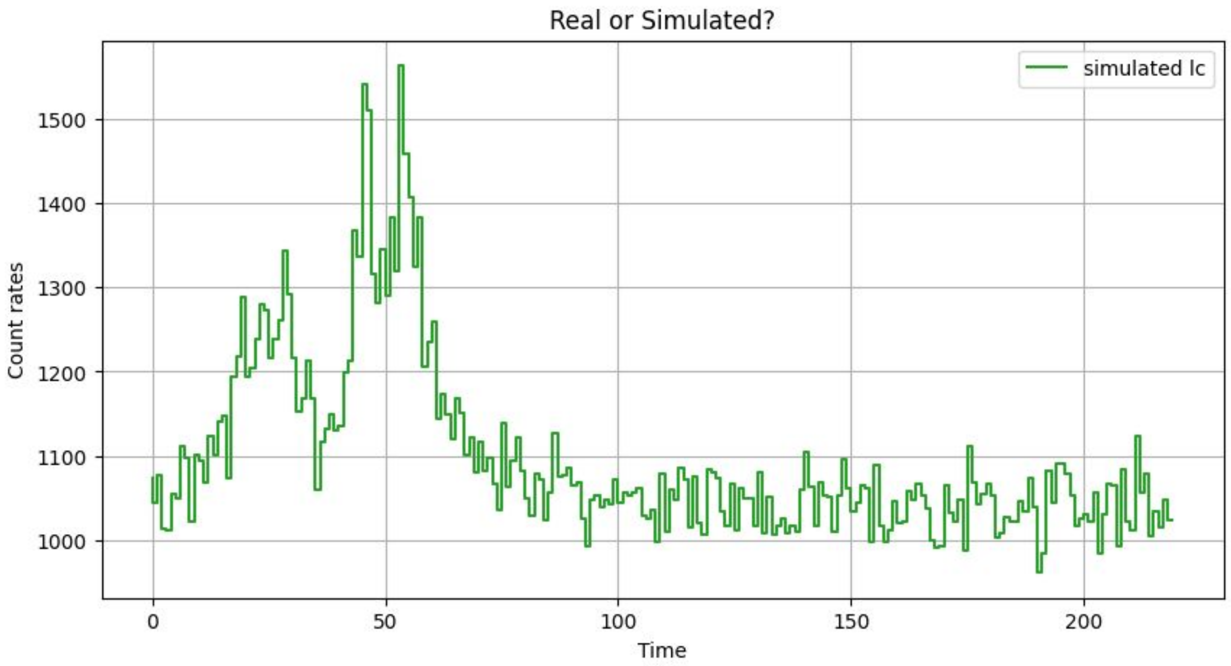


Simulated or real GRBs?

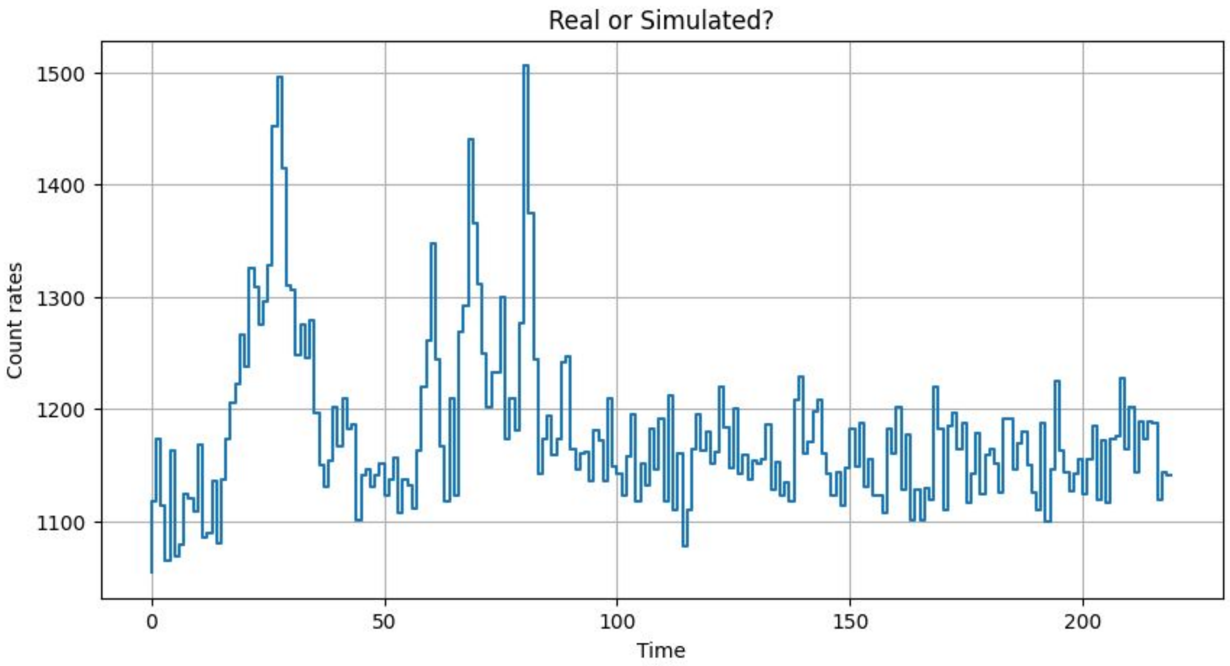
Real LC

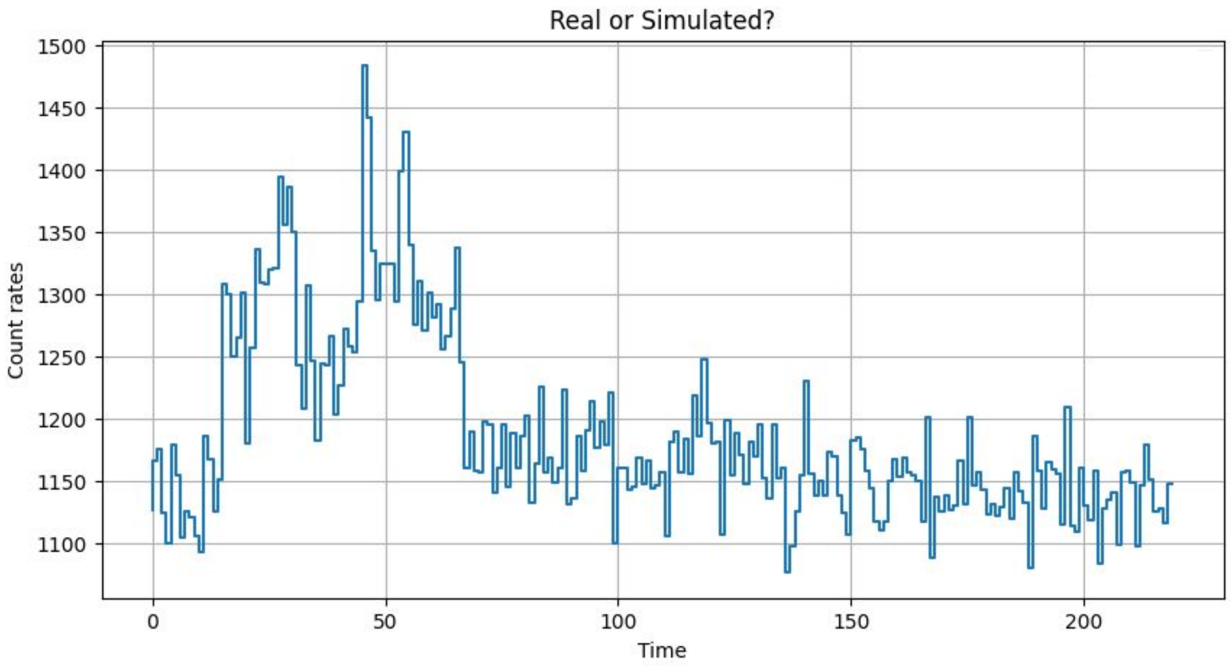


Simulated LC



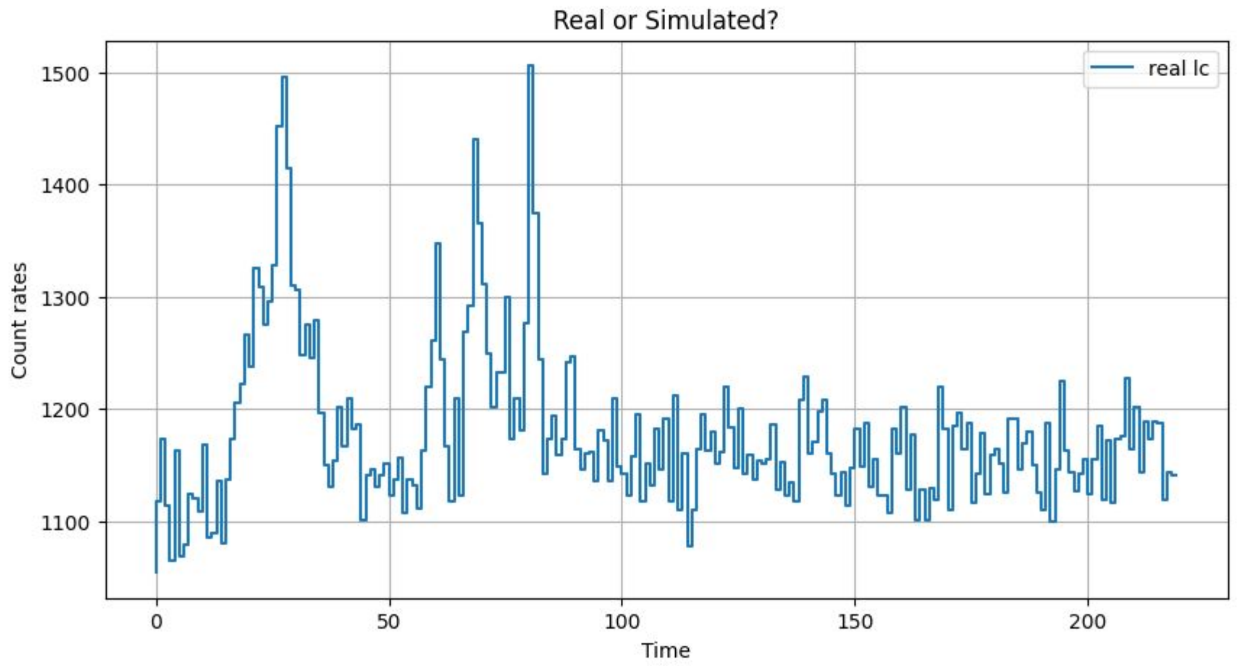
Simulated or real GRBs?



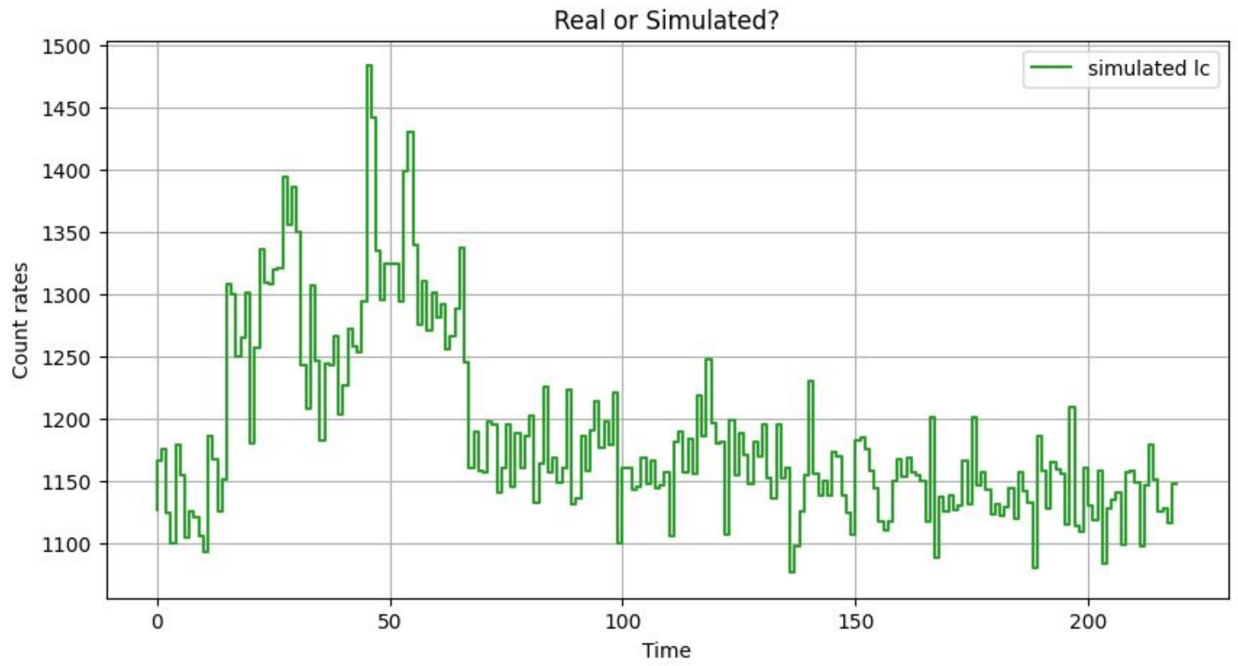


Simulated or real GRBs?

Real LC

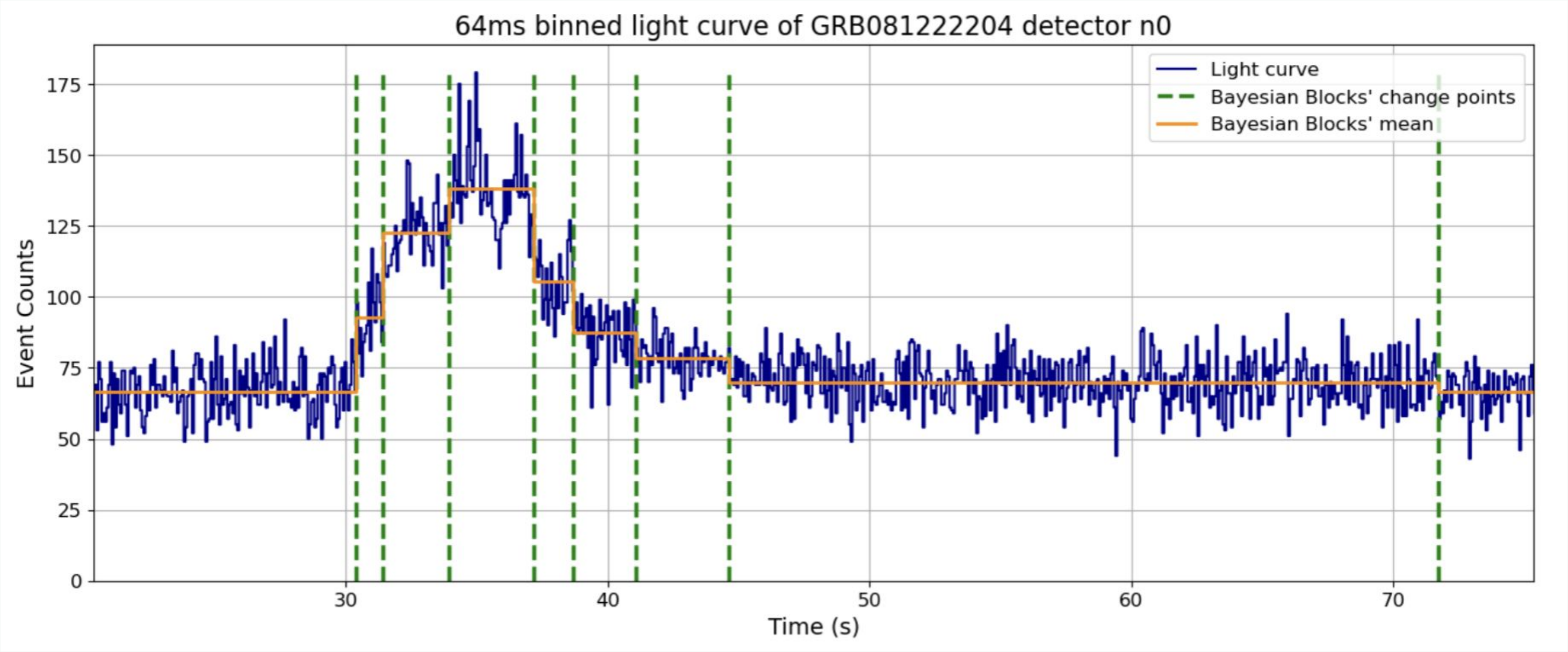


Simulated LC



Future works: Bayesian Blocks

Using **Bayesian Block** and **extracting features to characterize simulated data** compared to **real ones**



credits:
Massimiliano Zago

TO SUM UP...

This work presents a **novel DL model to generate synthetic GRBs' LCs**.

This model can be used to generate a **large dataset** to train and evaluate GRB detection methods.

Small modifications to the generative model **VAEGAN** to fit our task.

A huge effort was spent in the **data analysis** and **preprocessing** which allowed an improvement of the training dataset.

FUTURE WORKS

1. An evaluation methods based on **Bayesian blocks**
2. A **conditional version** of this model
3. A **probabilistic generator**, instead of generating mere LCs
4. **Combine this model with PINN** if a partial differential equation of GRB is found

THANKS FOR THE ATTENTION

Contacts: riccardo.falco@inaf.it