KRITTIKA SUMMER PROJECTS 2023

GRB Hunters: Faintest of the Brightest

Yash Salunkhe
Mentors: Gaurav Waratkar and Meghna Dixit

KRITTIKA SUMMER PROJECTS 2023

GRB Hunters: Faintest of the Brightest

Yash Salunkhe

Mentors: Gaurav Waratkar and Meghna Dixit

PUBLISHED BY KRITTIKA: THE ASTRONOMY CLUB OF IIT BOMBAY
Project Code Repository: KSP-GRB

Abstract

Gamma-ray bursts (GRBs) represent the most luminous and energetic explosions observed in the vast expanse of the universe. These cataclysmic events arise from the merger of neutron stars or the gravitational collapse of massive stars, emitting an initial burst of X-rays. Leveraging the capabilities of India's X-ray space telescope, AstroSat CZTI, over 500 GRBs have been successfully observed and documented. Furthermore, analysis has unveiled more than 50 previously unseen GRBs, establishing a remarkable discovery rate that suggests the potential existence of fainter GRBs within our data.

The primary objective of the project titled 'GRB Hunters' is to delve deeper into our extensive dataset and pursue the detection of these elusive, faint GRBs. Through rigorous examination and analysis, we aim to discern genuine GRBs from false candidates by employing quantitative measures of their detection characteristics. This methodology allows us to classify the identified events as either real GRBs or insignificant anomalies.

.

Contents

Part One

1. Gamma-Ray Bursts, AstroSat

1.1 Introduction

Gamma-Ray Bursts(GRBs) are high-energy explosions in distant galaxies lasting for times varying from milliseconds to hours. Most observed GRBs are supernovas(high-mass star implodes forming a black hole) or two colliding stars. These explosions are millions of light years away from the Earth and are very rare making it difficult to record them.

Figure 1.1: A picture of an exploding star creating a GRB

GRB light curves are incredibly complex and diversified. The duration of observable emission can vary from milliseconds to tens of minutes, there may be one peak or several, and individual peaks may be symmetric or exhibit fast brightening and very slow fading. No two gamma-ray burst light curves are alike; large variation is observed in almost every property.

1.2 AstroSat

AstroSat, India's first space observatory, was launched into low-Earth orbit on 28 September 2015. AstroSat carries four co-pointed instruments giving broadband coverage in optical,ultra-violet, soft X-ray and hard X-ray bands, as well as soft X-ray all-sky monitor.

In this project, we use the observations captured by the Cadmium Zinc Telluride Imager(CZTI) the wide-field instrument providing the hard X-ray coverage for AstroSat. It detects energy levels in the range of 20-200+ keV.

Here's a schematic of the CZTI on AstroSat:

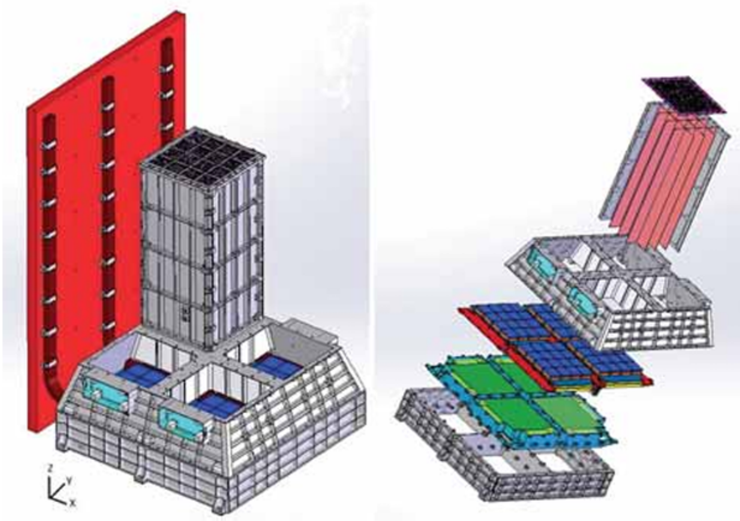


Figure 1.2: The blue squares attached to the red module are the CZTIs(right image). The black film on top is the coded aperture mask)

2. CZTI Pipeline

2.1 Working of the Pipeline

The data received from the CZTI payload passes through a series of steps in a processing pipeline. Three major data levels have been designated for long-term storage. These are:

1. Level 0 – This is the raw data received from satellite telemetry, which is segregated by instrument, along with auxiliary data. This data is archived internally and not distributed for public use.
2. Level 1- This is reorganized raw data, written in FITS format for Astronomical use. All auxiliary information necessary for further processing of this data is collated at this level and packed along with the respective science data.
3. Level 2 – This data contains standard science products derived from Level 1 data. Level 2 data is also in FITS format and is available for science use, with the same lock-in criteria and release mechanism as the Level 1 data.

2.2 Operating the Pipeline

Here are a series of steps to use the pipeline after you have completed installing it on your system:

1. Get the sample data or data of any GRB you want to observe. You will need the `bc.evt` and `livetime.fits` from the `modeM0` folder in Level 2, `.mkf` file from Level 1 and the `mkf_treshholds.txt` file.
2. Run `cztigen`: Inputs would be the paths to the `.evt`, `.mkf` and `mkf_treshholds.txt` file. Store the output preferably in the same directory.
3. Run `cztdatasel`: Inputs are `bc.evt`, `cztgtigenoutput` (gti).
4. Run `cztpixclean`: Inputs are `bc_datasel.evt`, livetime file, `par_det_count_thresh=100`, `par_pix_count_thresh=1000`
5. Run `cztevtclean`: Input `quad_pixclean.evt`
6. Run `cztflagbadpix`: Input the pixclean output (`quad_badpix`).

7. Run `cztbindata`: Input `quad_clean.evt`, mkfile, badpix output (badpix file), livetime file, binsize = 1 sec. Filling an energy range often gave segmentation errors so keep it blank.

After successfully performing these steps, the user will generate lc and pha files which will be used to observe GRBs!

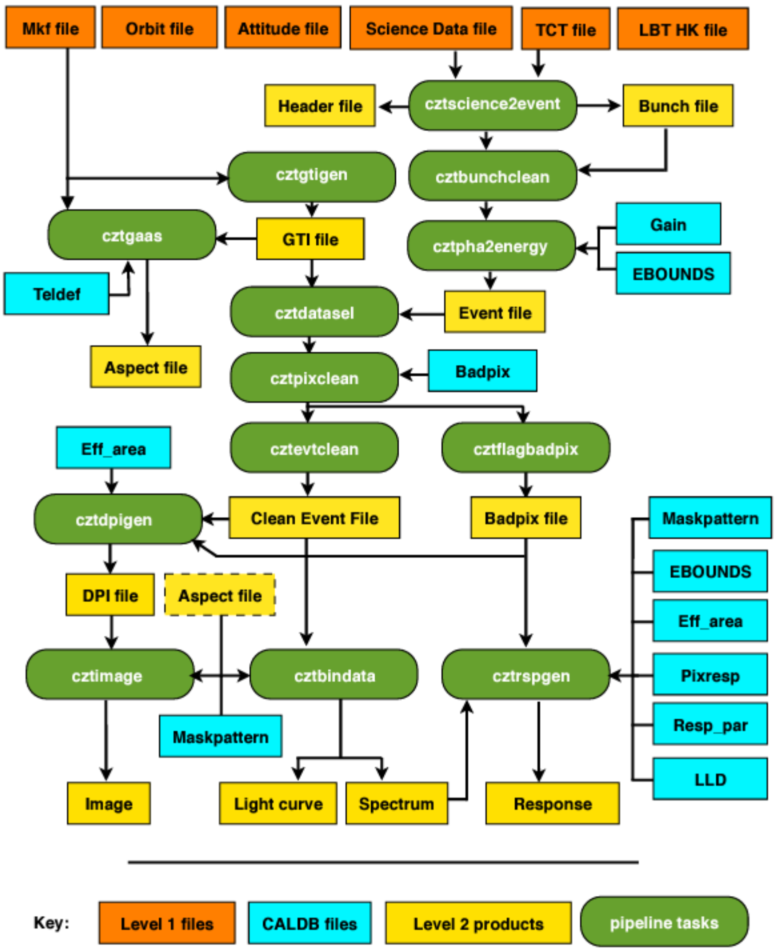


Figure 2.1: A schematic which shows all internal operations of the pipeline.

Part Two

3. Lightcurves

3.1 Using lc files

After obtaining the lc files, to access the data in it, use the `fits` module in the `astropy` package. You can import it in your Jupyter Notebook by running the command -

```
from astropy.io import fits
m1=fits.open('filename.lc')   #Replace with appropriate filename
```

Checking the contents and plotting the count rate vs time will give you the required lightcurve.

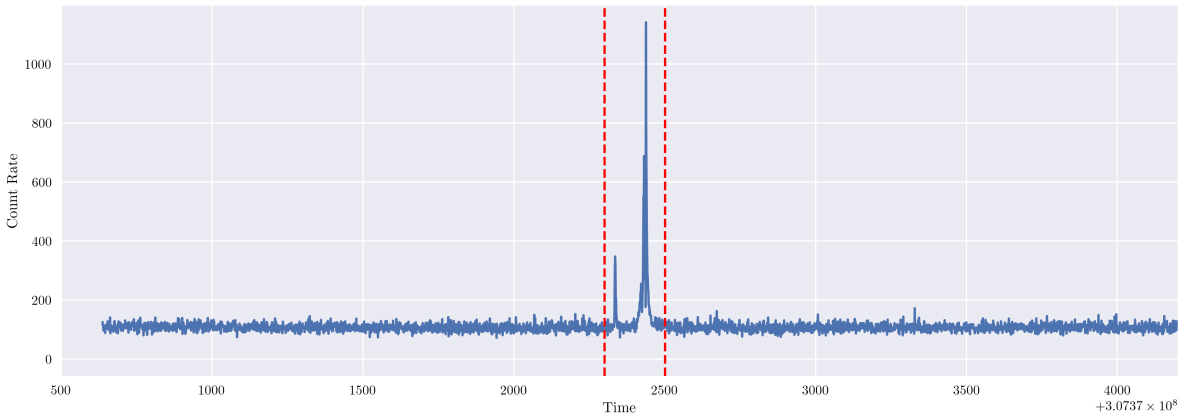


Figure 3.1: Lightcurve obtained from using the sample data (Quad 0)

The dip and flat line in the later half of the graph is due to the South Atlantic Anomaly. It is an area where Earth's inner Van Allen radiation belt comes closest to Earth's surface. This leads to an increased flux of energetic particles in this region and exposes orbiting satellites to higher-than-usual levels of ionizing radiation.

One can also plot spectral plots using the data in the `pha` files. Plotting the counts vs channels from the `pha` files will give the required spectral curve.

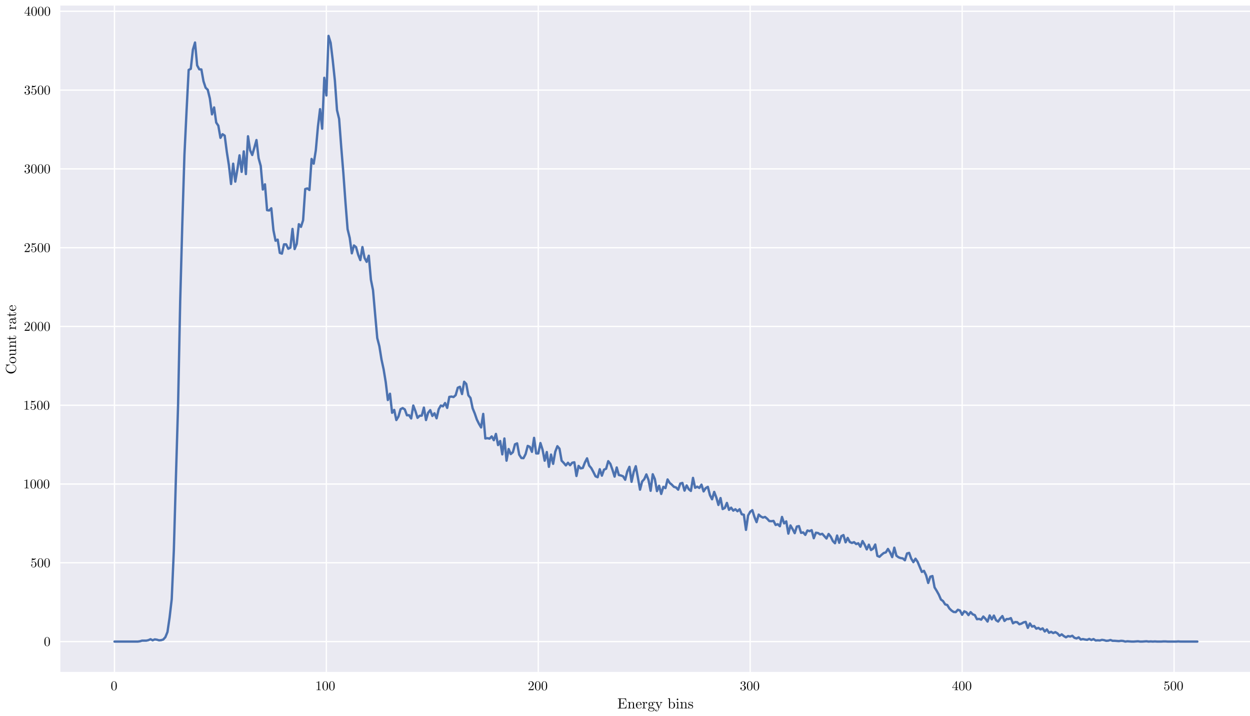


Figure 3.2: Spectral curve from the sample data

The x-axis of the spectral curve are the energy bins, 20-200 keV divided into 512 bins. Hence, this shows the distribution of energy in the data.

4. Detrending lightcurves

4.1 Introduction

Removal of trends from a time series data set is called detrending. Detrending time series provides us with normal cyclic time series data as it removes non-cyclic trends. This process tells about the factors affecting the time series data and finds important factors about it.

4.2 Methods to detrend lightcurves

4.2.1 Median filter

Median filtering is a popular technique used for detrending signals. It is a nonlinear filtering method that aims to remove or reduce the effects of outliers and short-term fluctuations while preserving the underlying trend or low-frequency components of the signal. The main idea behind median filtering is to replace each data point

in the signal with the median value within a specified neighborhood or window. By using the median instead of the mean or other statistical measures, the filtering process becomes robust against extreme values or outliers that can distort the trend estimation. Here are steps to perform median filtering:

1. Choose a suitable window size or neighborhood that determines the number of adjacent samples to consider.
2. Start by positioning the window at the beginning of the signal.
3. Compute the median value within the window.
4. Replace the original data point at the center of the window with the difference between the original data and median in that window.
5. Slide the window to the next position along the signal and repeat steps 3 and 4 until the entire signal has been processed.

6. The resulting signal is the detrended version, where the trend components and short-term fluctuations have been reduced or removed while preserving the underlying trend information.

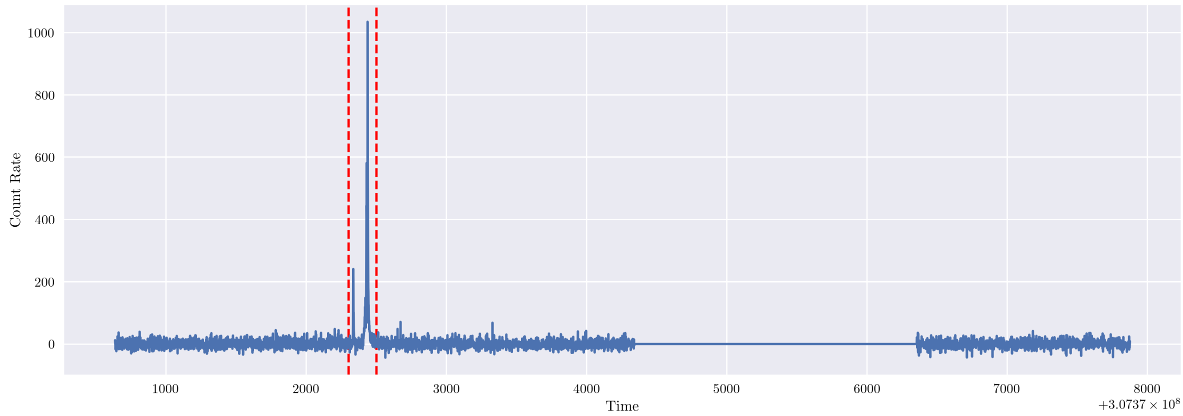


Figure 4.1: Detrended lightcurve with window size = 5

4.2.2 Savitzky–Golay(Savgol) filter

A Savitzky–Golay filter is a filter that can be applied to a set of data points for the purpose of smoothing the data, that is, to increase the precision of the data without distorting the signal tendency. This is achieved, in a process known as convolution, by fitting successive sub-sets of adjacent data points with a low-degree polynomial by the method of linear least squares.
One can implement the Savgol filter in python by using `savgol_filter` from the `scipy.signal` library.

Figure 4.2: Savgol filtered signal with a 3rd order polyomial

4.3 Signal to Noise ratio(SNR)

Signal-to-noise ratio, often written S/N or SNR, is a measure of the strength of the desired signal relative to background noise (undesired signal). S/N can be determined by using a fixed formula that compares the two levels and returns the ratio, which shows whether the noise level is impacting the desired signal.
The ratio is typically expressed as a single numeric value in decibels (dB). The ratio can be zero, a positive number or a negative number. A signal-to-noise ratio over 0 dB indicates that the signal level is greater than the noise level. The higher the ratio, the better the signal quality.

4.3.1 Traditional method

The traditional method to calculate SNR involves taking the mean of the signal remove the mean of the background from it and divide it by the noise. Mathematically, it turns out to be:

$$SNR = \frac{\text{Mean}(signal) - \text{Mean}(background)}{\text{Standard dev.}(background)} \quad (4.1)$$

To convert these values into units of dB, take logarithm(base 10) of this fraction and multiply 20 to it.

The SNR for the four quadrants of the sample data using this method turns out to be:

Quadrant number	SNR (in dB)
Quad 0	15.395
Quad 1	12.845
Quad 2	20.167
Quad 3	18.378

4.3.2 RMS method

The root-mean square (RMS) of a variable is the value of the square root of the sum of the squares of the stacking variable values divided by the number of values. Here, we take the RMS value of the noise. Mathematically, the SNR would look like-

$$SNR = \frac{\text{Mean}(signal) - \text{Mean}(background)}{\text{RMS}(background)} \quad (4.2)$$

The SNR for the four quadrants of the sample data using this method turns out to be:

Quadrant number	SNR (in dB)
Quad 0	19.661
Quad 1	17.893
Quad 2	24.577
Quad 3	21.123

4.3.3 Gaussian modeling of data

Plotting a histogram of the count rate against its frequency gives a Gaussian-like looking histogram. Using `curve_fit` from `scipy` gives the best curve-fit for the Gaussian distribution function.

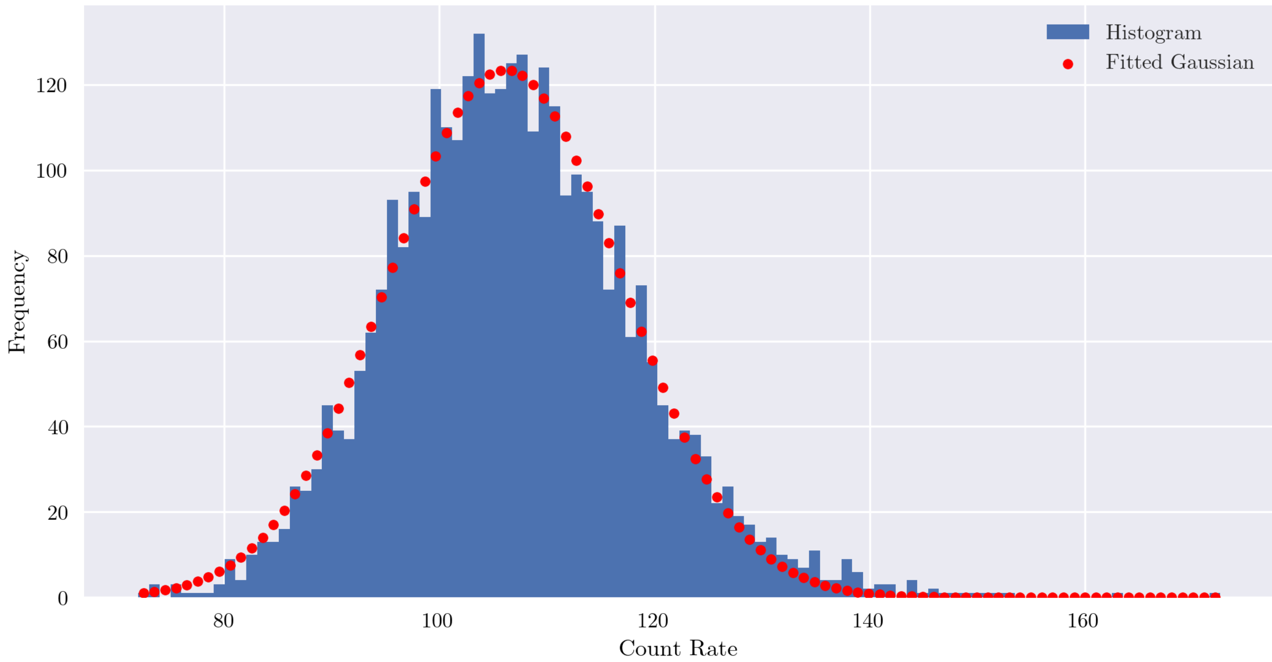


Figure 4.3: Gaussian distribution fitted to a histogram

To calculate SNR, we take the mean, and standard deviation values of the fitted Gaussian, and we consider mean + 3*standard deviation as the threshold for our background as 99.73% of the data lies behind this threshold.
Mathematically, the SNR for the 4 quadrants would be:

$$SNR = \frac{Maximum(signal)}{\sigma} \tag{4.3}$$

The SNR for the four quadrants of the sample data using this method turns out to be:

Quadrant number	SNR (in dB)
Quad 0	30.931
Quad 1	29.020
Quad 2	35.259
Quad 3	31.622

5. Timebins

5.1 Introduction

The role of timebins in light curves is to aggregate and organize the observational data into discrete time intervals or bins. Each timebin represents a specific time span during which observations are combined, and the data within that interval are summarized or averaged. Timebins are particularly useful when dealing with unevenly spaced or irregularly sampled observational data.

5.2 Effect of different timebins

For the fainter GRB200803A, I took timebins of 0.1s, 0.5s, 1s, 10s (default is 1s) and applied the above explained SNR algorithms to study the effect of timebins on the SNR values. The variation was as seen:

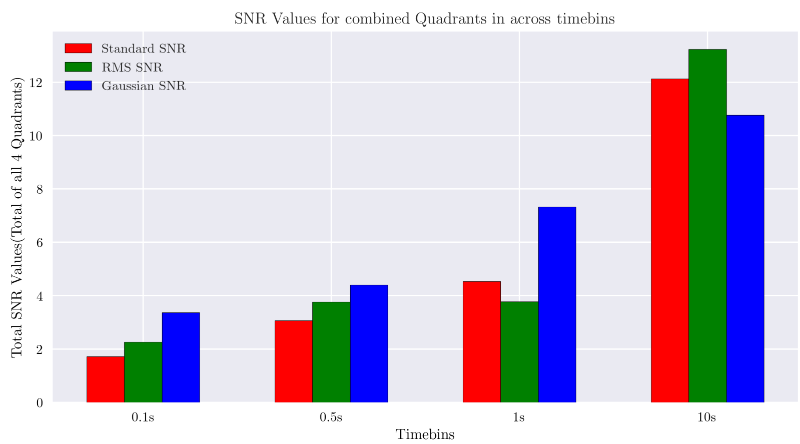


Figure 5.1: Variation of SNR with changing timebins

The Signal-to-Noise Ratio (SNR) of a light curve exhibits variability with changing timebins. Smaller timebins result in higher temporal resolution, capturing rapid fluctuations in brightness, and potentially revealing finer details of the object's behavior. This finer resolution often reduces the SNR since it enhances the impact of noise and reduces the detection of real signals. However, using extremely small timebins may lead to sparse data points, limiting the accuracy of the overall SNR estimation. Conversely, wider timebins offer better smoothing, reducing noise impact but sacrificing temporal details. Choosing an appropriate timebin width is crucial in balancing temporal resolution and SNR reliability in light curve analysis.

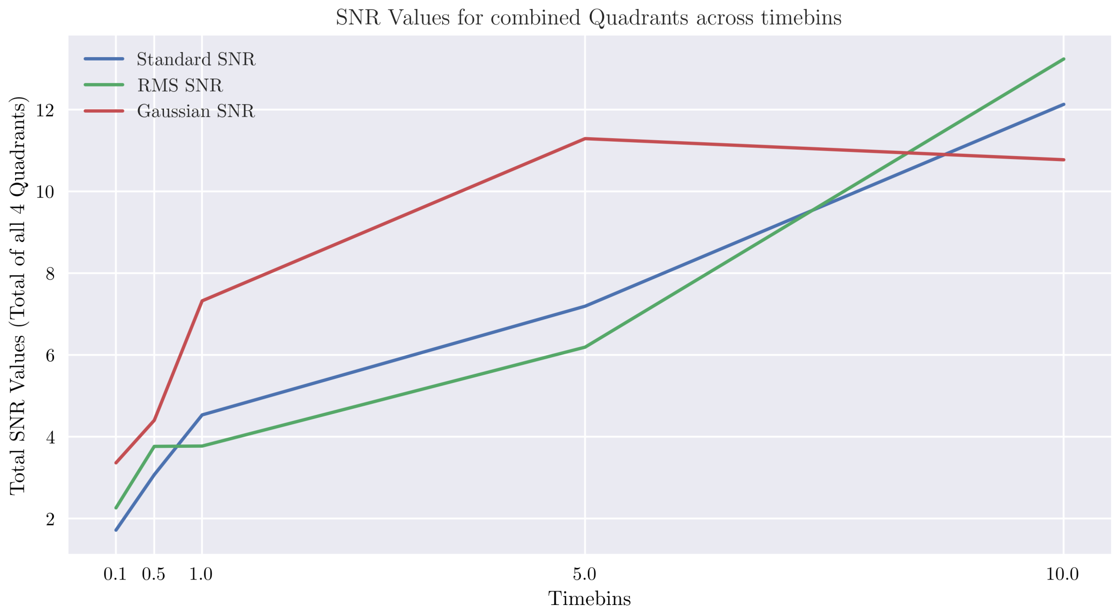
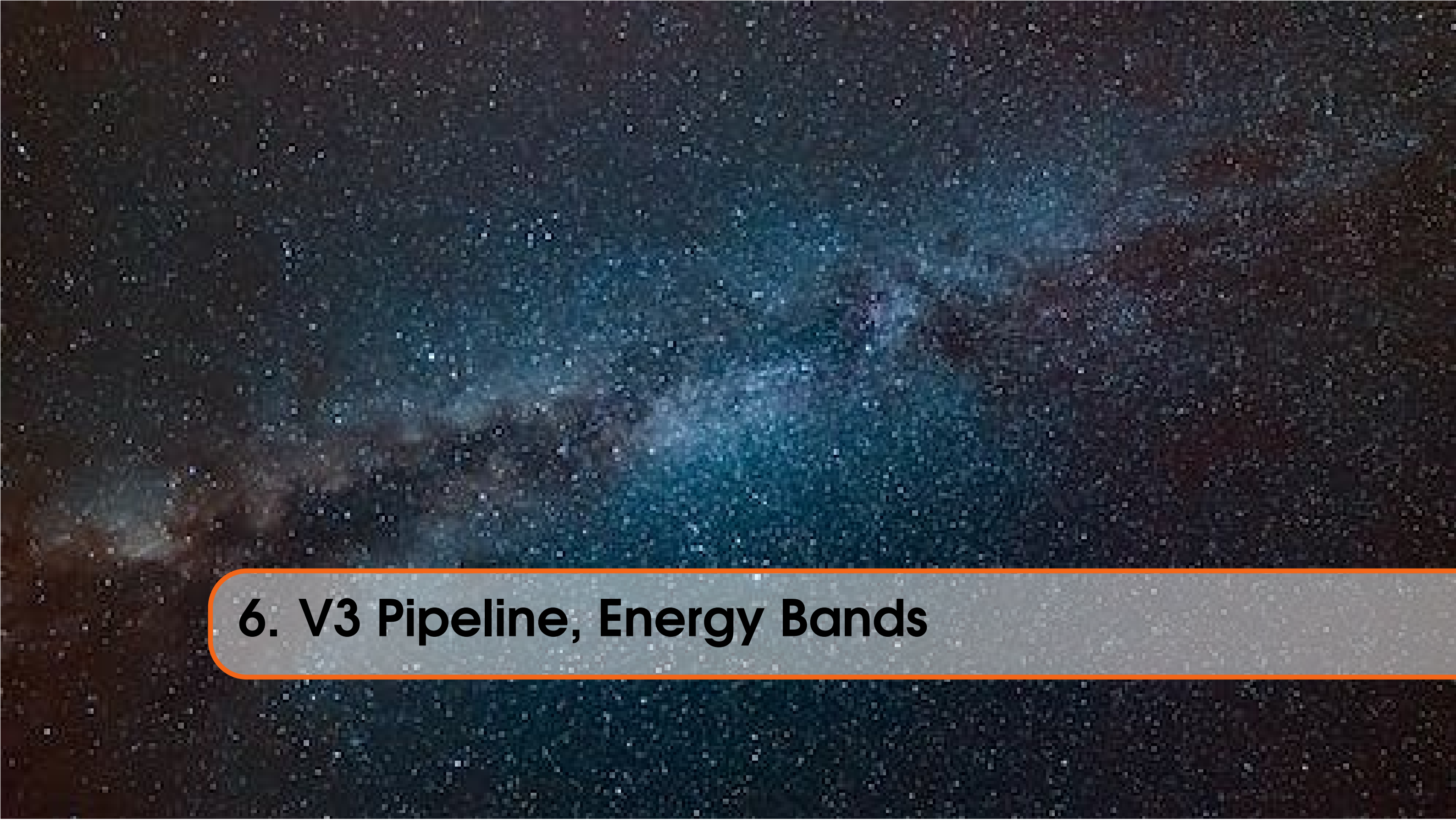


Figure 5.2: Variation of SNR with changing timebins

However, sometimes a bin size which is not the maximum may have the highest SNR of all(quite frequent in faint GRBs). In this case this SNR would be considered as the optimal SNR and all computations would be perfomed on this SNR



6. V3 Pipeline, Energy Bands

6.1 Introduction

The V3 version of the pipeline provides stronger noise cleaning algorithms and the option to view lightcurves in energy bands. In V2, the lightcurves were seen in the entire energy range but in V3 the user has the option to split the energy bands.

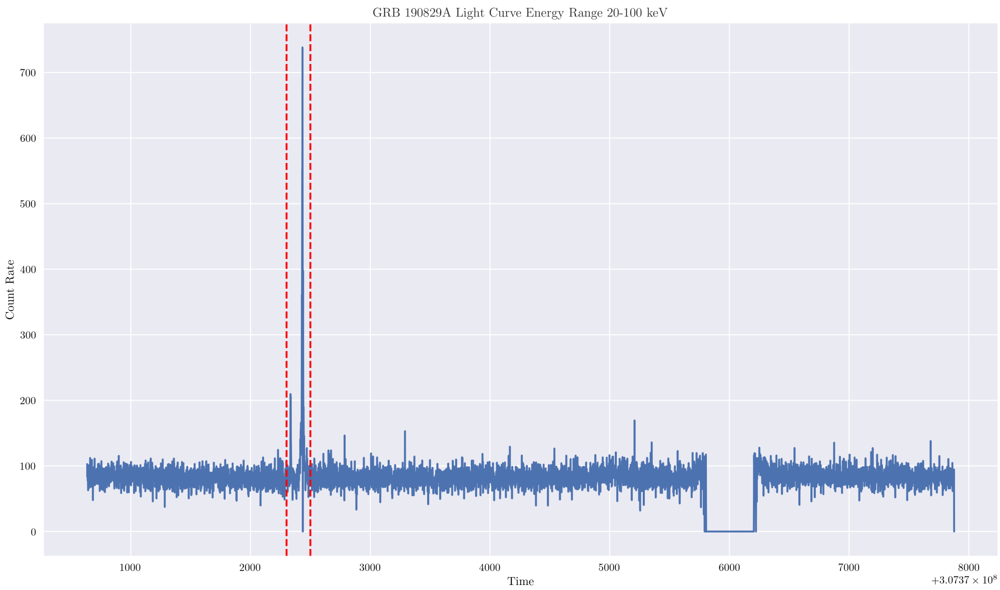


Figure 6.1: 20-100 keV energy band for GRB190829A

20 - 100 keV band and 100 - 200 keV bands refer to specific energy ranges within the electromagnetic spectrum. GRBs emit high-energy photons, and the lightcurves represent the time evolution of the GRB's brightness at different energy

levels. The choice of energy bands is crucial for studying the properties of GRBs and understanding their underlying physical processes.

20 - 100 keV band: This energy band typically represents the low-energy part of the GRB spectrum. It includes photons with energies between 20 and 100 kiloelectron volts (keV). Lightcurves in this band provide information about the less energetic emissions of the GRB and are often used to study the initial prompt emission phase of the burst. This phase is believed to be related to the internal processes of the GRB central engine.

100 - 200 keV band: This energy band represents the higher-energy part of the GRB spectrum. It includes photons with energies between 100 and 200 keV. Lightcurves in this band capture the higher-energy emissions of the GRB. Analyzing this band is important as it provides insights into the high-energy processes occurring during the later stages of the burst.

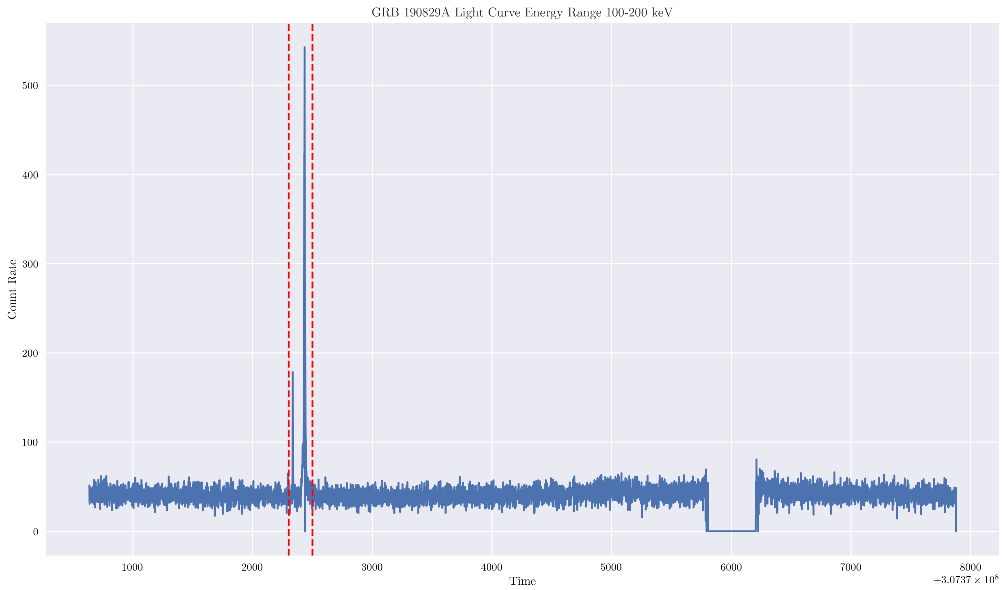


Figure 6.2: 100-200 keV energy band for GRB190829A

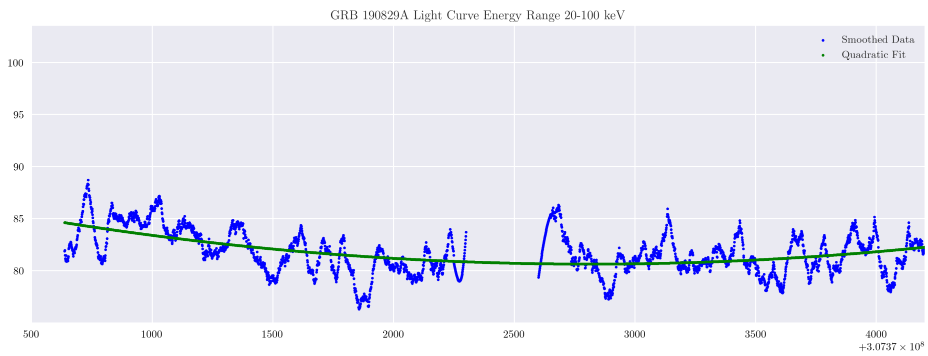


Figure 6.3: A quadratic fit to SavGol filtered 20-100 keV data

The SavGol filter effectively reduces noise and fluctuations, revealing underlying trends and features. Subsequently, subtracting a quadratic function from the original lightcurve helps detrend the data, removing long-term variations or systematic effects, thereby isolating short-term behaviors. This approach is valuable in identifying intrinsic signal variations, aiding in the accurate characterization of GRB properties and associated physical processes.

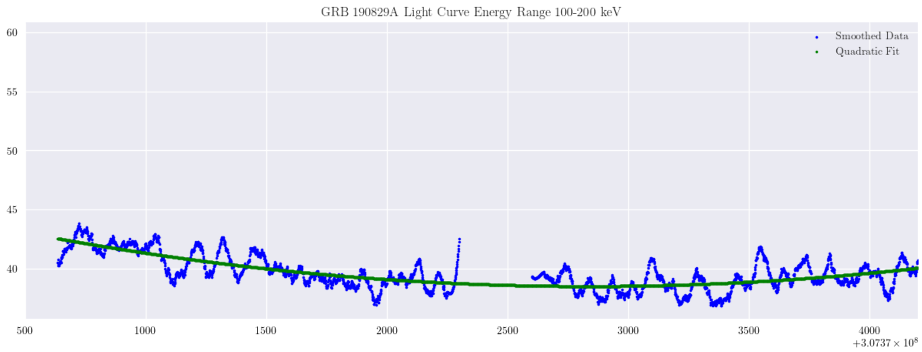


Figure 6.4: A quadratic fit to SavGol filtered 100-200 keV data

6.2 Particle or GRB?

Using two energy bands to classify a signal as a gamma-ray burst (GRB) or a particle event is a common practice in high-energy astrophysics. The key lies in understanding the characteristic properties of each type of event in different energy ranges. Gamma-ray bursts emit intense and brief bursts of high-energy photons. The signal from a GRB is expected to be concentrated in high-energy bands often with a well-defined temporal structure corresponding to the burst's prompt emission phase.
On the other hand, particle events, such as background cosmic-ray interactions or instrumental noise, may produce a relatively constant or slowly varying signal across a single energy band. They generally have a single point in the spectrum and don't have a continuous distribution.

We were given 2 possible GRBs - GRB210519A and GRB210709A and were tasked to identify whether these were particles or GRBs. Upon analysis and applying the above-mentioned algorithms, I measured the SNR of both these lightcurves at their trigger time at varying energy bands.

My Conclusion
GRB210709A - Is a GRB
GRB210519A - Is a particle (SNR dipped at higher energy bands)

References

- Wikipedia - GRBs
- Research paper on CZTI on Astrosat
- Research paper explaining the working of the CZTI Pipeline
- Site to download the pipeline and sample data
- Wikipedia- South Atlantic Anomaly
- Wikipedia- Savitzky-Golay (Savgol) filter
- Wikipedia- Median filtering
- Wikipedia- Signal-to-Noise ratio
- Wikipedia- RMS
- Wikipedia- Gaussian distribution
- CZTI V3 Pipeline Guide