# Exploring the utilization of elastic data binning for time-series representation for Tomo-e light curve data

Phungtua-eng Thanapol*, 西川 侑志, 山本 泰生

静岡大学・情報学部

**木曽シュミットシンポジウム2023**

2023.05.30

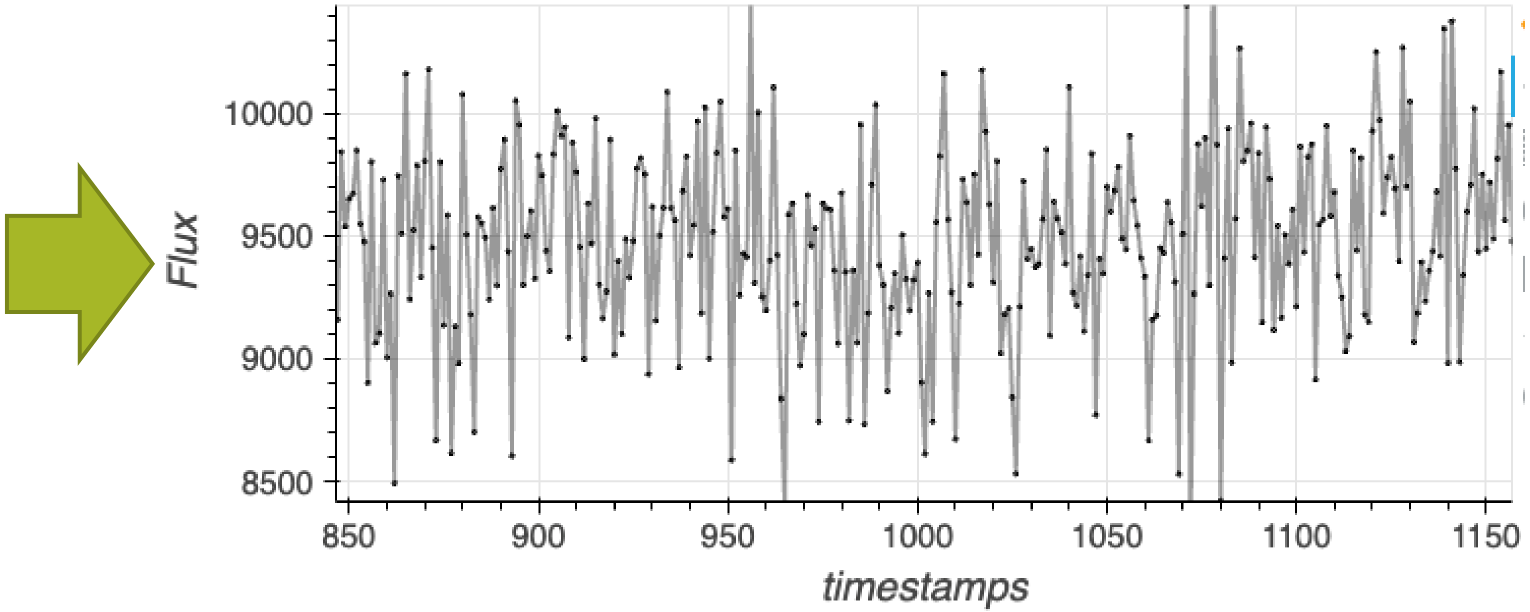
# Outline

- Background on time series data mining
- The importance of **right** sketching
- Elastic data binning
    - Sketching results using EBinning
    - Transient pattern detection results
- Appendix: Matrix Profile/STL with lightcurve data

# Time series data

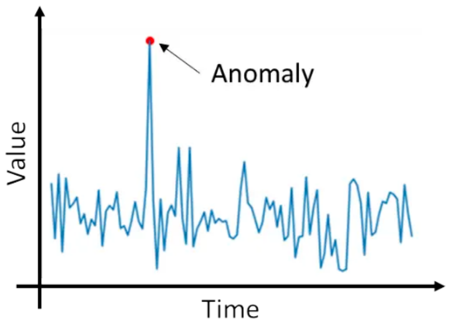
- A time series data is a collection of observations made sequentially in time.
- A lightcurve is a type of time series data

9845.0
9540.0
9656.0
9677.0
9852.0
9550.0
…
…
9343.0
9602.0
9709.0
10021.0
9439.0

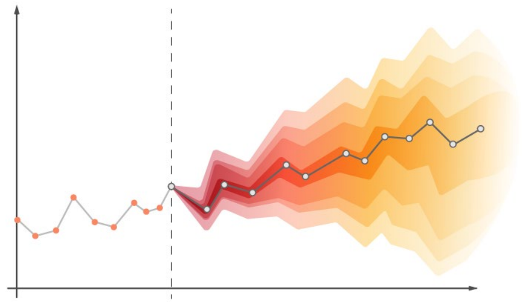


# Time series data mining

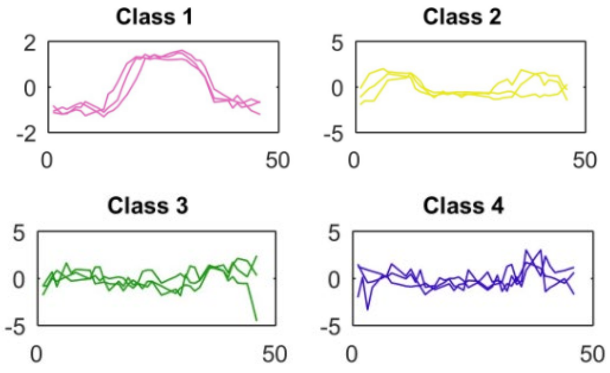
- What can we do with time series data?
  - Classification
  - Anomaly detection
  - Forecasting
  - Etc…
- What are the problems with time series data?
  - Noise
  - Data overload makes it difficult to extract meaningful or insights.




Anomaly detection



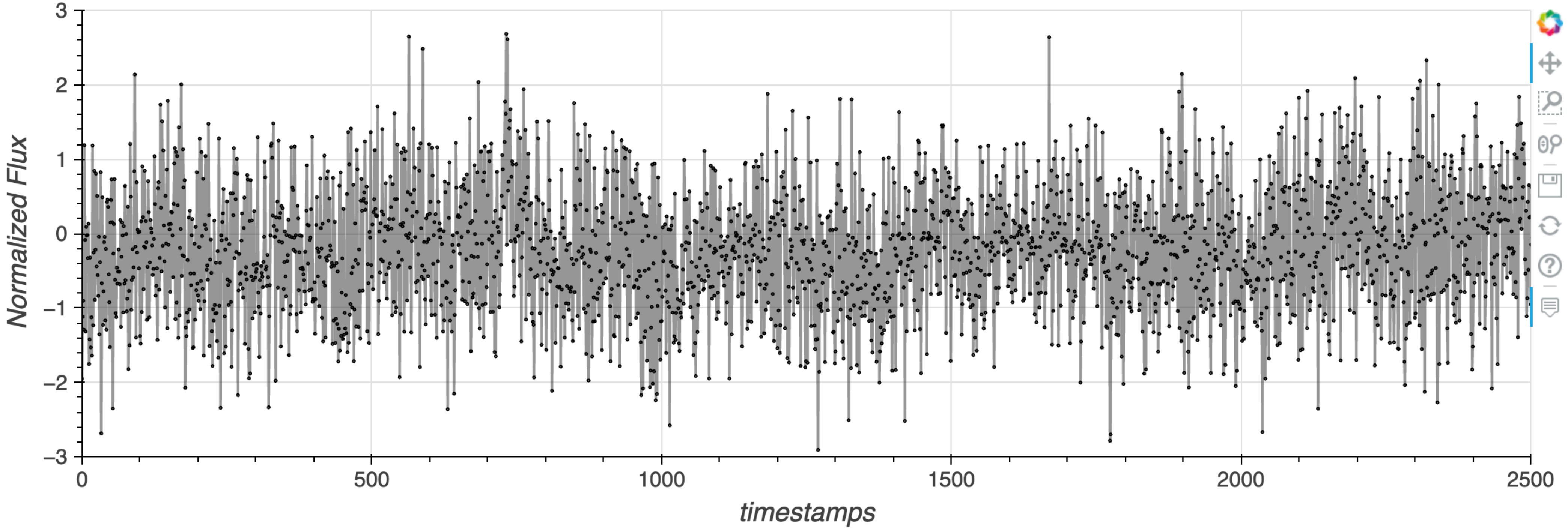
Forecasting



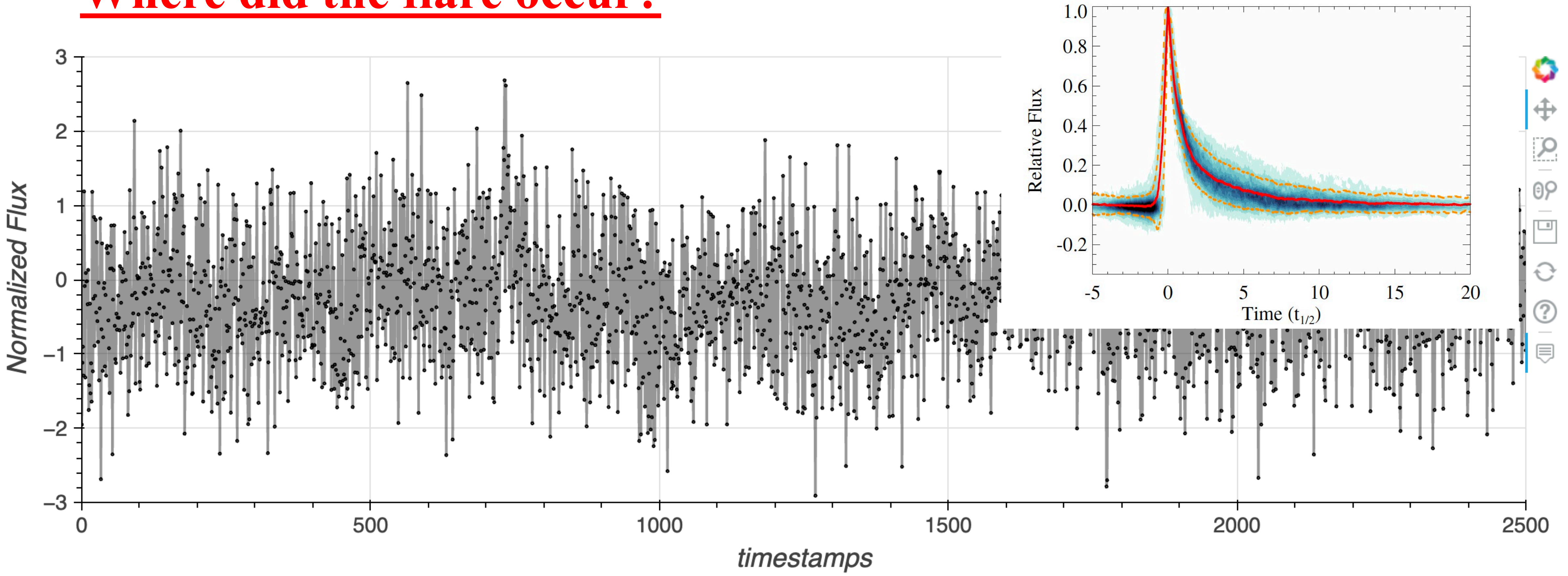

Classification

# The importance of **right** sketching

- Example of ligthcurve data from Tomo-e Gozen.

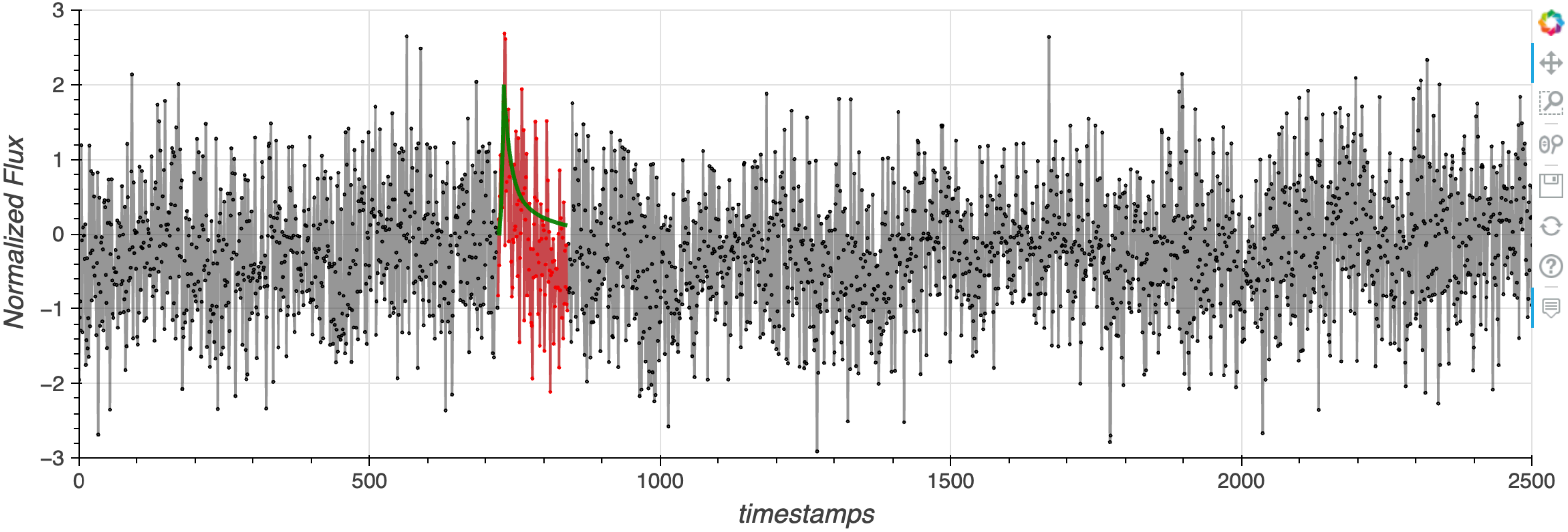


- We injected a synthetic flare pattern into this lightcurve.
- The synthetic flare pattern is based on the flare model of Kepler flare [1].
- **Where did the flare occur?**

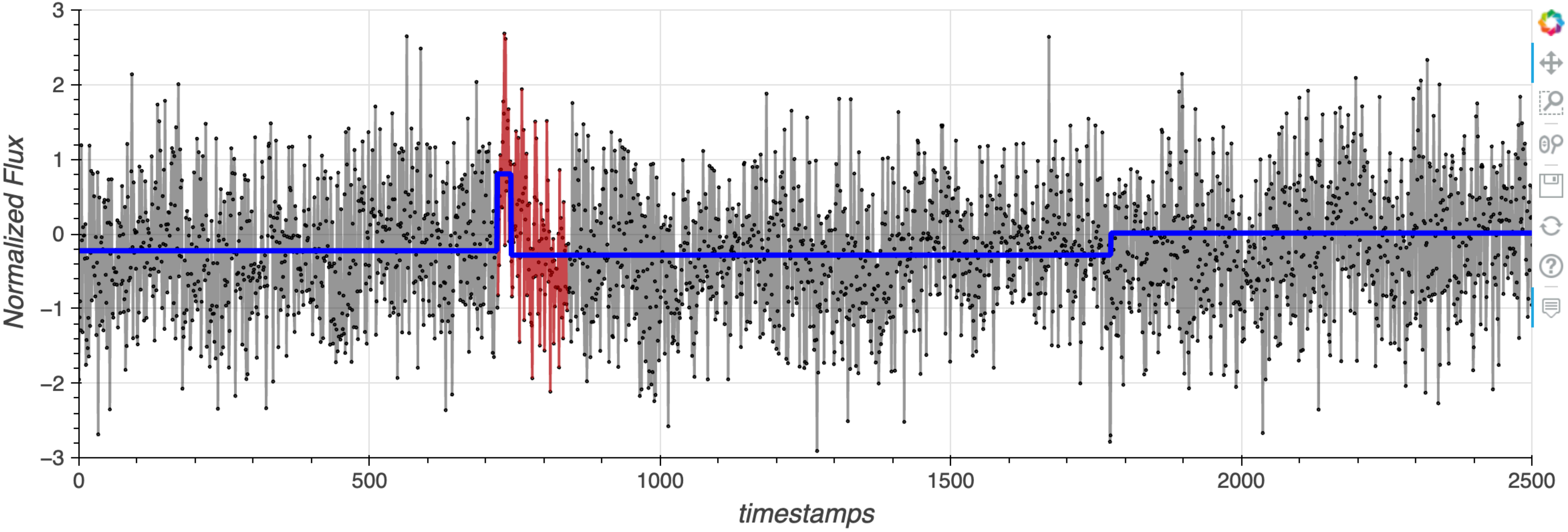



[1] "**Kepler Flares. II. The Temporal Morphology of White-light Flares on GJ 1243**" **:** James R. A. Davenport *et al* 2014 *ApJ* **797** 122

- The red highlight represents the synthetic flare occurring.
- What happen if we use right sketching?



- The blue line represents sketching using our proposed method [2].
- We can quickly identify when flare occurring.




[2] T. Phungtua-eng, Y. Yamamoto, and S. Sako. 2023. Elastic Data Binning for Transient Pattern Analysis in Time-Domain Astrophysics. ACM SAC '23, 342 – 349. : **(Online version will appear by July 2023.)**

# Solution : time series sketching

- What can we do for solving problems with time series data?
  - Use time series sketching!
- The advantage of time series sketching:
  - Faster processing: Sketches allow for quicker computations and analysis compared to raw data.
  - Key patterns and features: Sketches capture essential characteristics of time series data.

# Outline

- Background on time series data mining
- The importance of **right** sketching
- Elastic data binning
  - Sketching results using EBinning
  - Transient pattern detection results
- Appendix: Matrix Profile/STL with lightcuver data

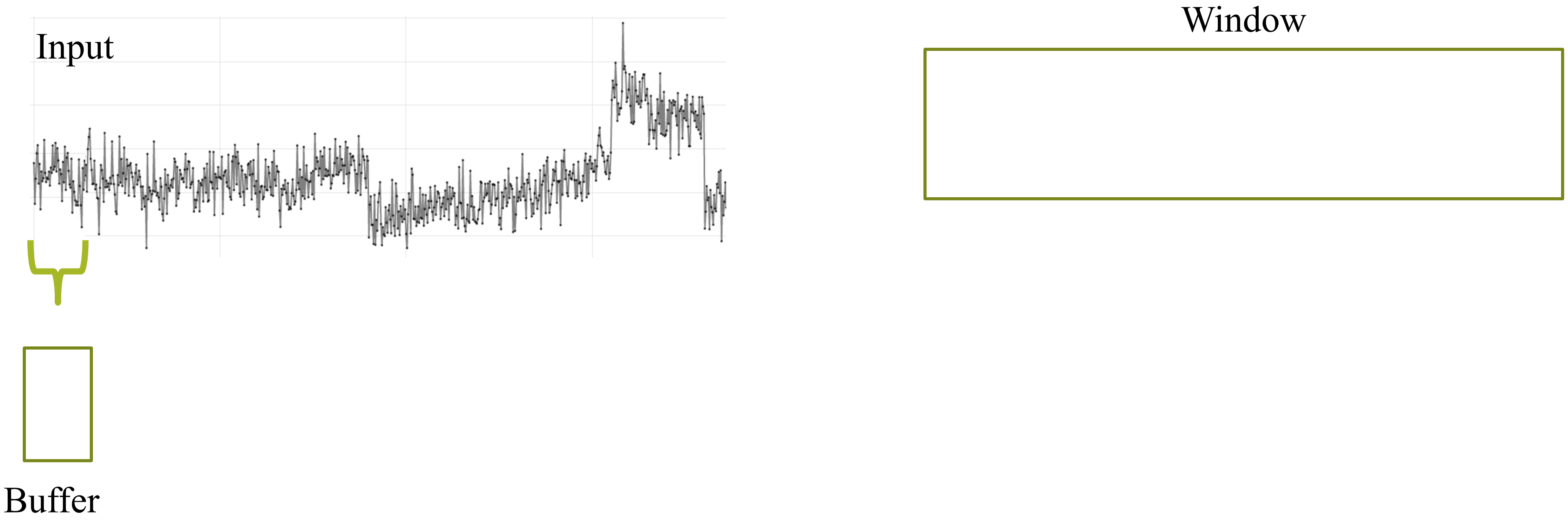
# Elastic data binning (EBinning)

- EBinning sketches lightcurve by partitioning it into bins of varying sizes and summarizing each bin with its mean.

- The highlight properties of the Ebinning:
  - It does not require any training process.
  - It has low-time complexity.

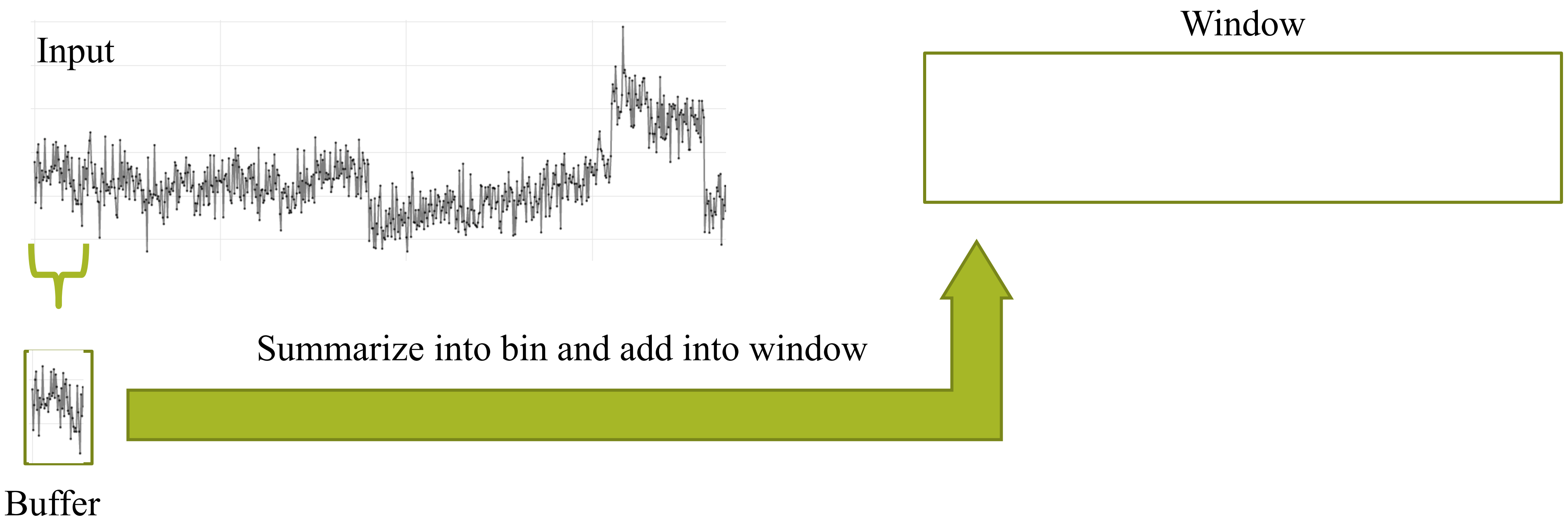
# Step 1: create empty bin

- Creates an empty buffer and adds subsequences to it.
- Once the buffer is full, it is summarized into bins and added to the window.



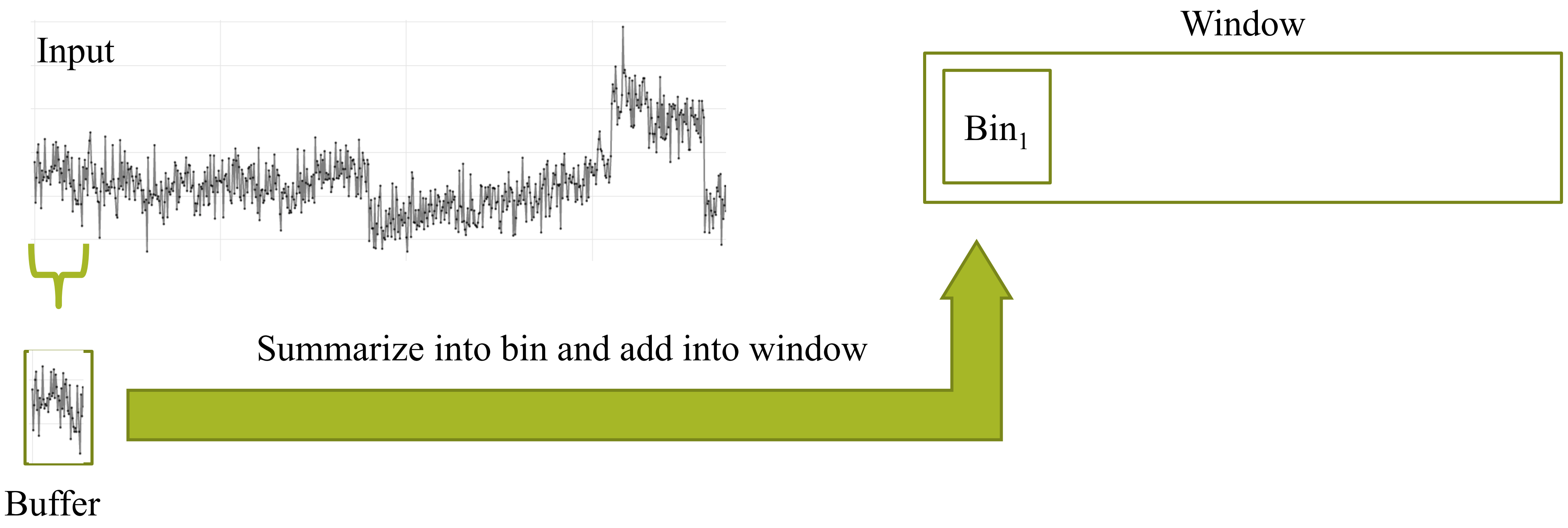
# Step 1: create empty bin

- Creates an empty buffer and adds subsequences to it.
- Once the buffer is full, it is summarized into bins and added to the window.



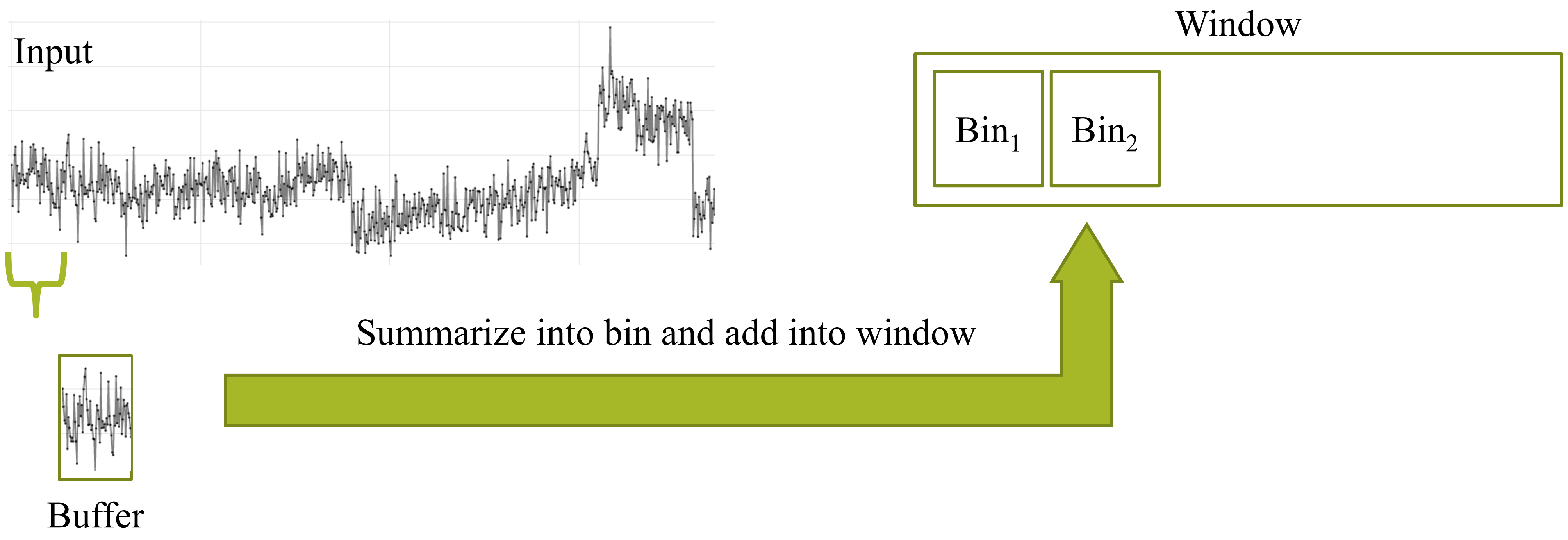
# Step 1: create empty bin

- Creates an empty buffer and adds subsequences to it.
- Once the buffer is full, it is summarized into bins and added to the window.



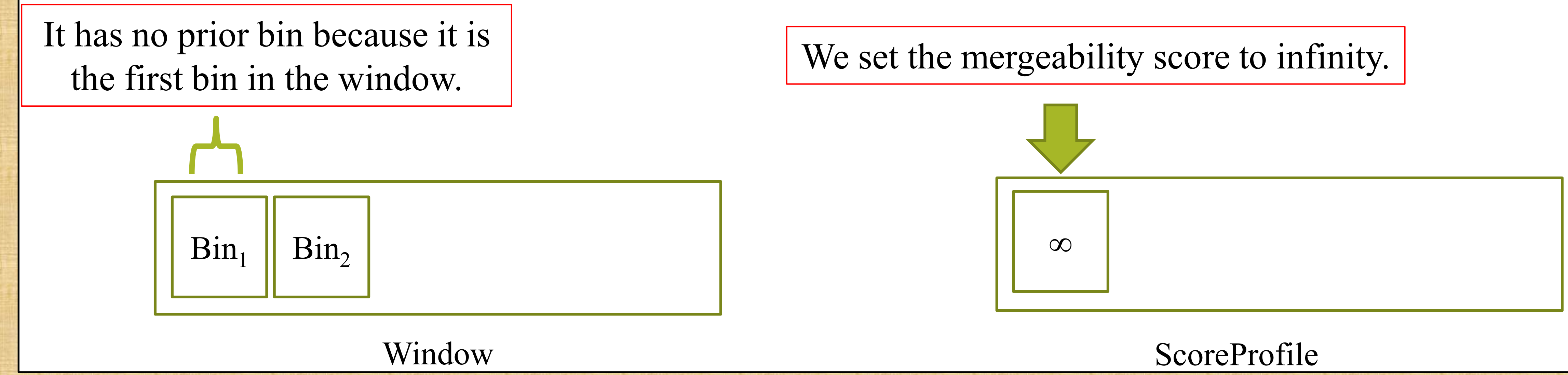
# Step 1: create empty bin

- Creates an empty buffer and adds subsequences to it.
- Once the buffer is full, it is summarized into bins and added to the window.



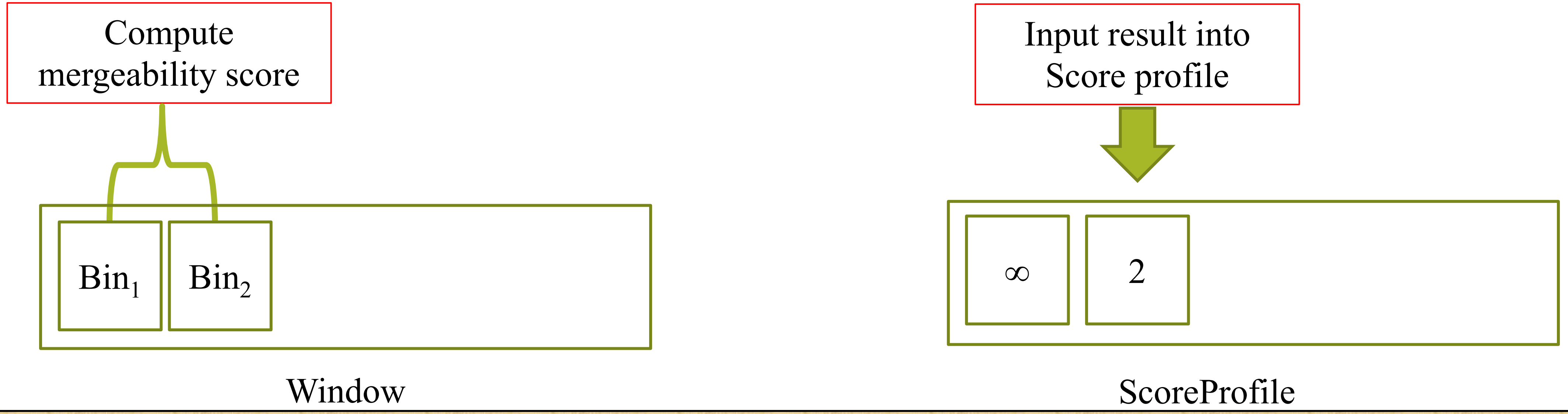
# Step 2 mergeability score computation

- We compute mergeability scores ($k$) between two neighboring bins in the window.
  - If $k$ is closest to zero, the neighboring bins have the same feature.
  - If $k$ is farthest from zero, the neighboring bins do not have the same feature.



# Step 2 mergeability score computation

- We compute mergeability scores ($k$) between two neighboring bins in the window.
  - If $k$ is closest to zero, the neighboring bins have the same feature.
  - If $k$ is farthest from zero, the neighboring bins do not have the same feature.



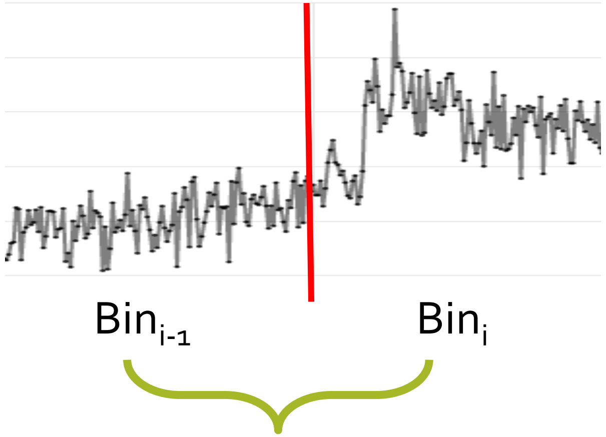
# Mergeability scores of EBinning

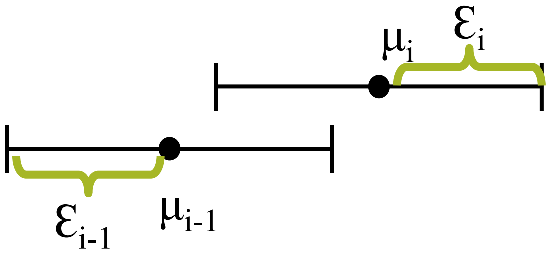
- Hoeffding’s inequality helps identify mergeable bins based on their mean values' boundaries (ε).

$$\varepsilon = \sqrt{\frac{(b-a)^2}{2n} \log \frac{2}{\delta}}$$

($\delta$ is user-defined error probabilty)

- Thereafter, we test two bins for distributed equality.



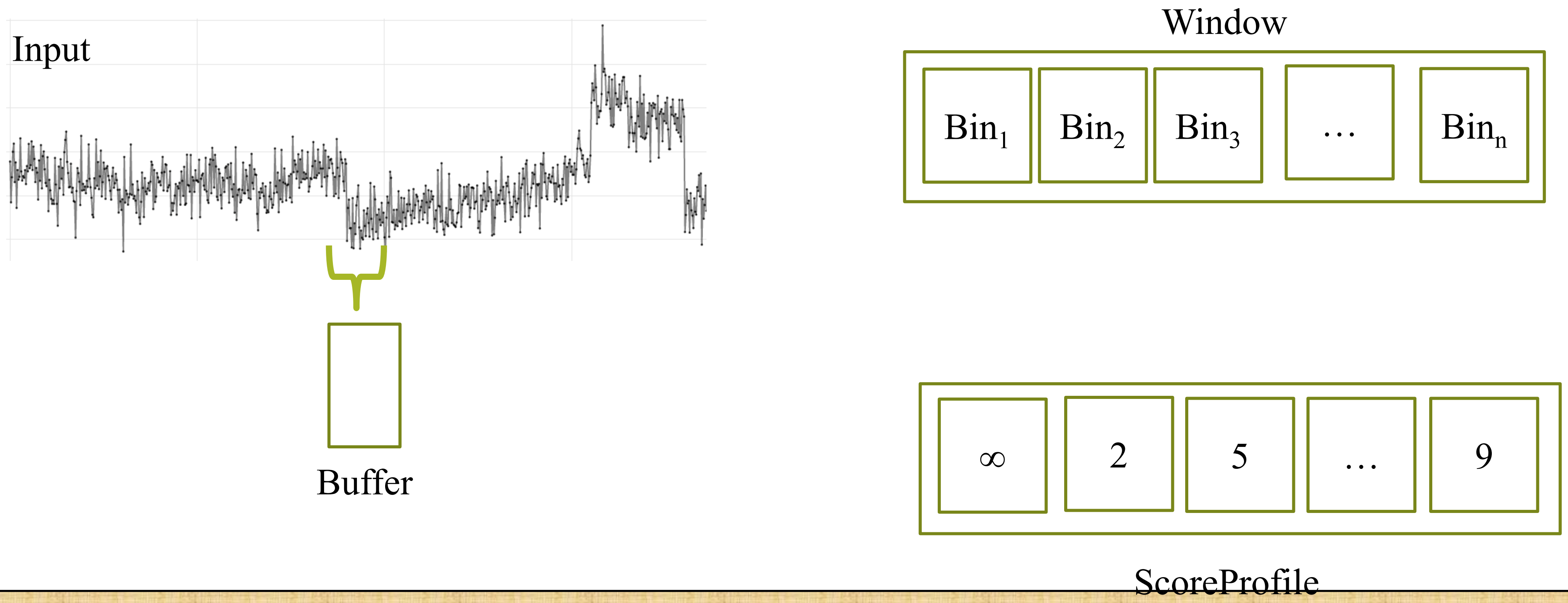




Check whether two boundaries belong to the same distribution or not.

$$k = \frac{|\mu_i - \mu_{i-1}|}{\min(\varepsilon_i, \varepsilon_{i-1})}$$

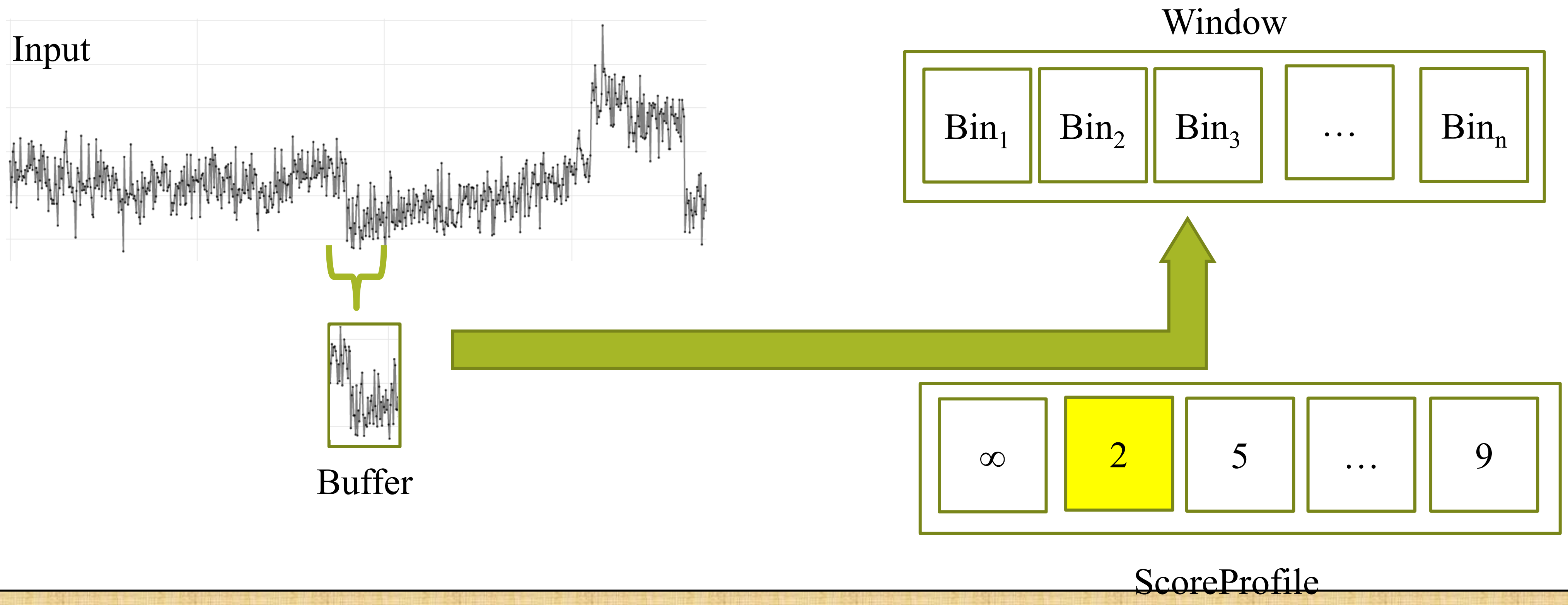
# Step 2 mergeability score computation

- Add bins and compute mergeability score until the window is full.



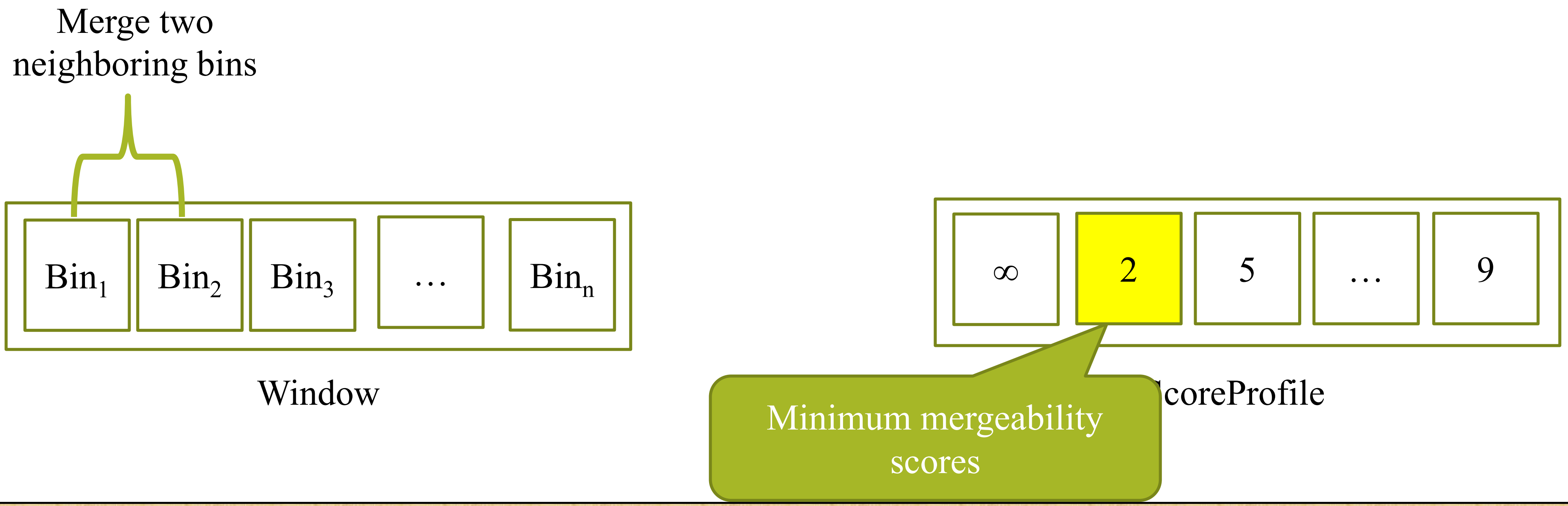
# Step 2 mergeability score computation

- Then select the index with the minimum mergeability score for merging.



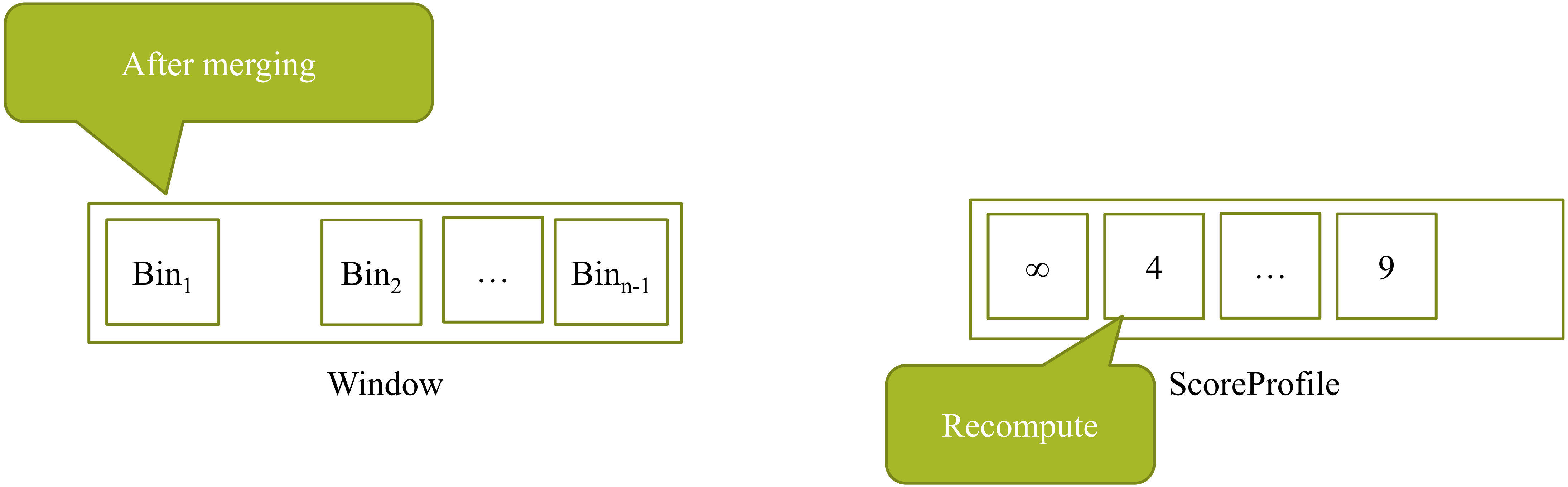
# Step 3 : Merging

- We merge bins with minimum mergeability scores when the window is full.

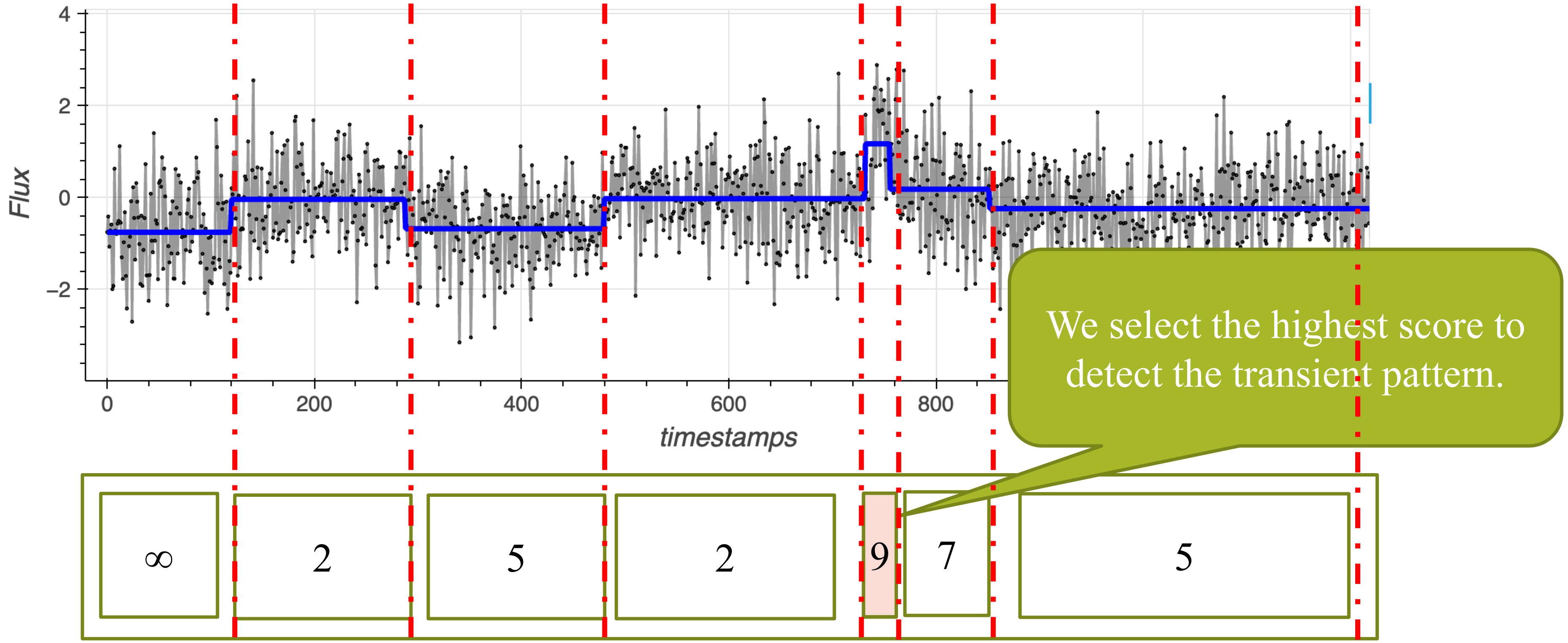


# Step 3 : Merging

- We merge bins with minimum mergeability scores when the window is full.
- Do the Step1 until the end of input.

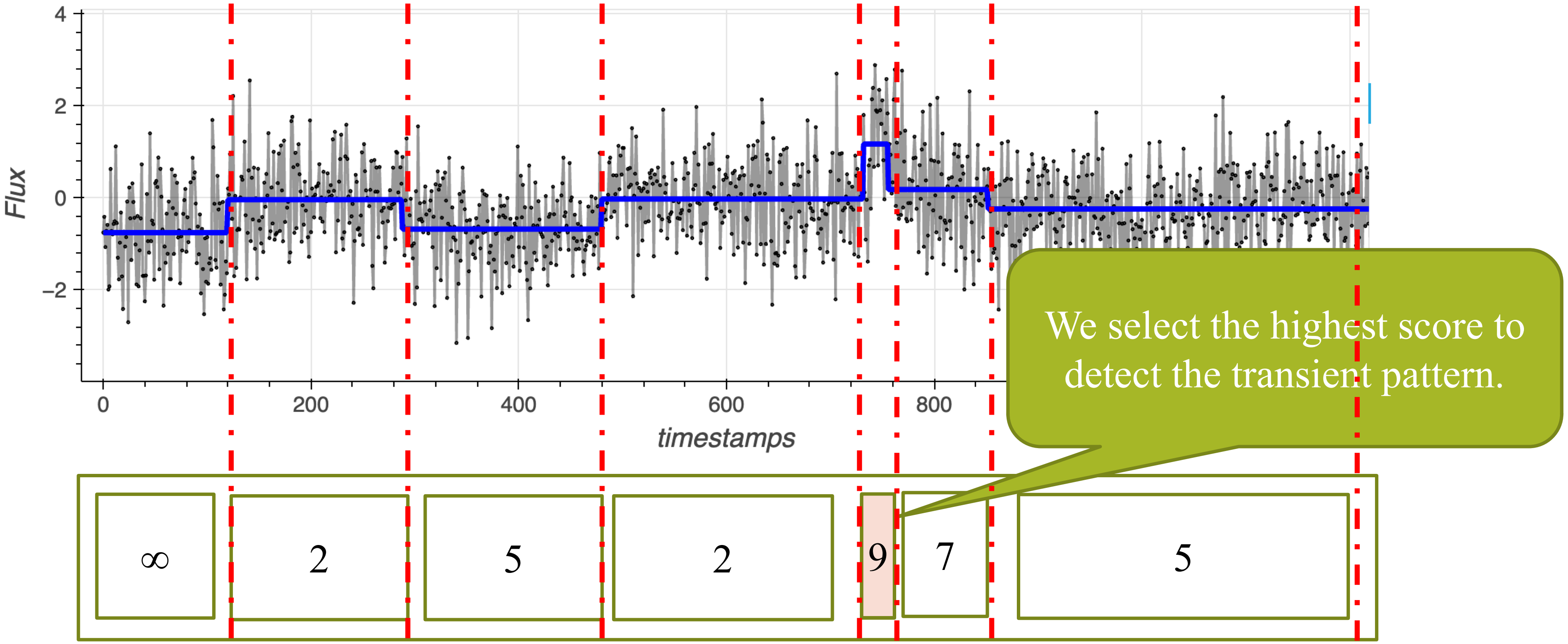


# How to read ScoreProfile?




ScoreProfile

# How to read ScoreProfile?

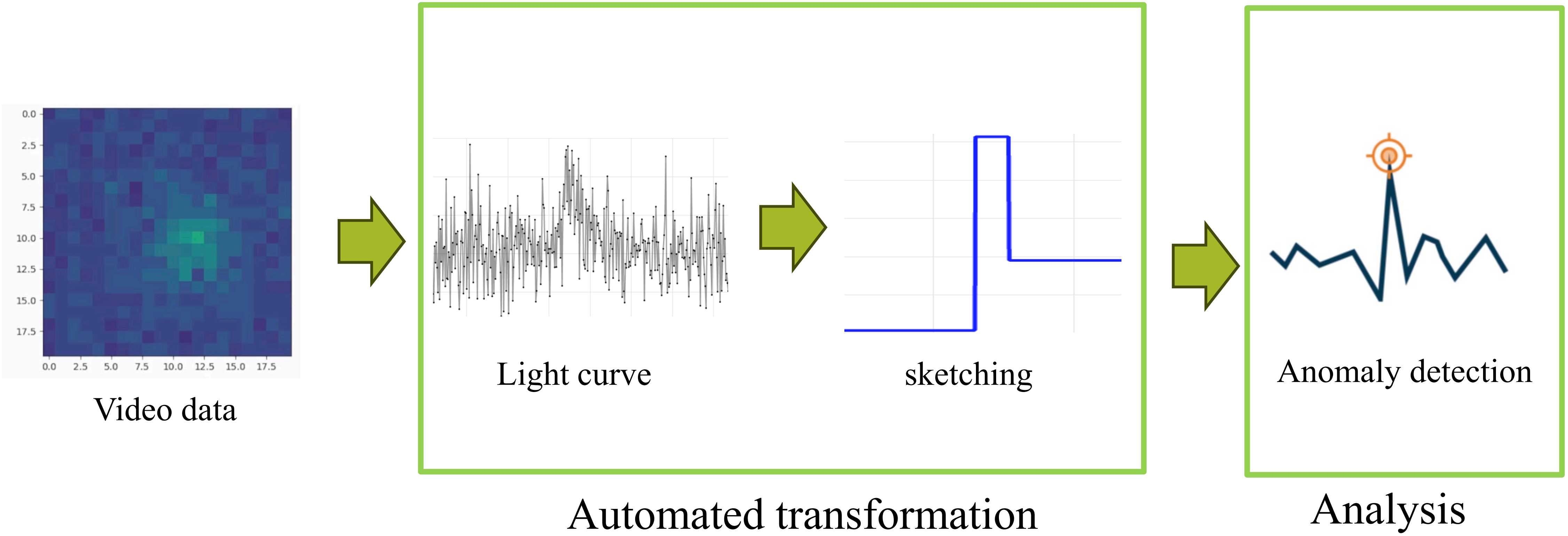



ScoreProfile

# Elastic data binning (EBinning)

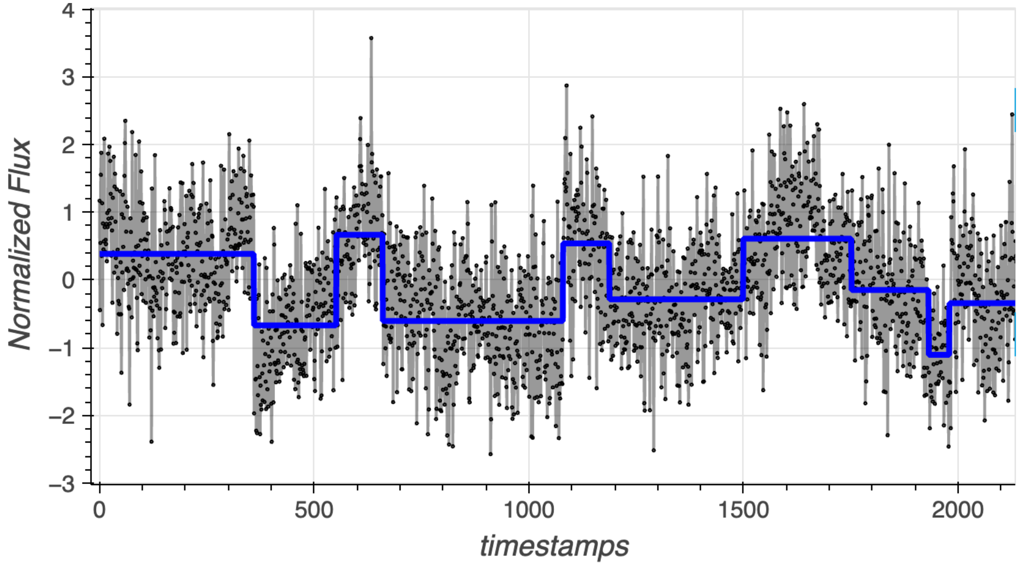
- Adjusting bin size based on specific time series characteristics.
- How to adjust bin for skecthing?
  - We will partition when we found the distribution changing between two bins.
  - We will merge when we found two bins are the same distribution.

# Utilization with **EBinning**

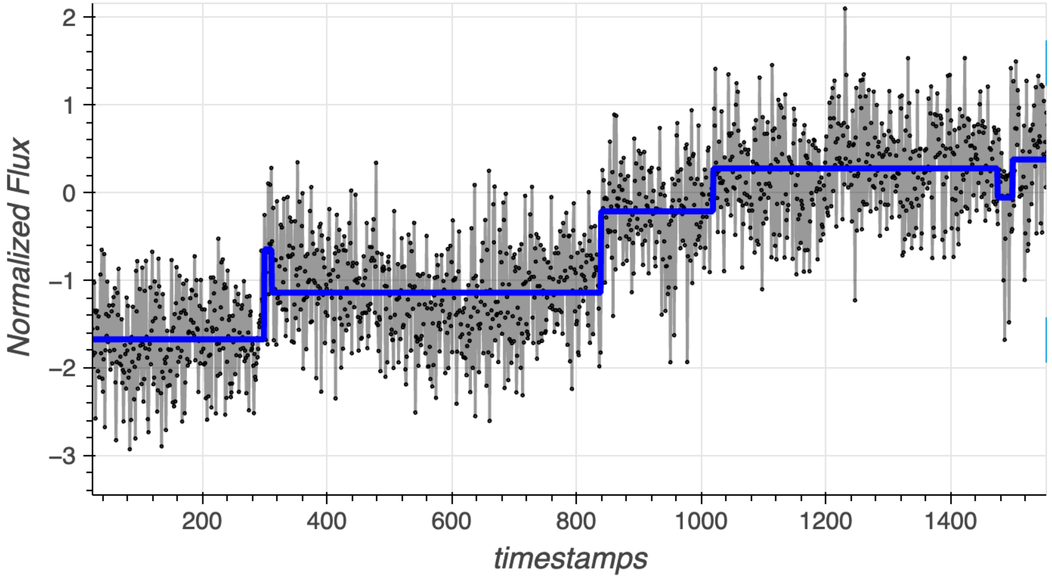


# Sketching results using EBinning

- Our sketching can emphasize the essential characteristics of the data.
  - Atmospheric opacity
  - Component of nearby stars



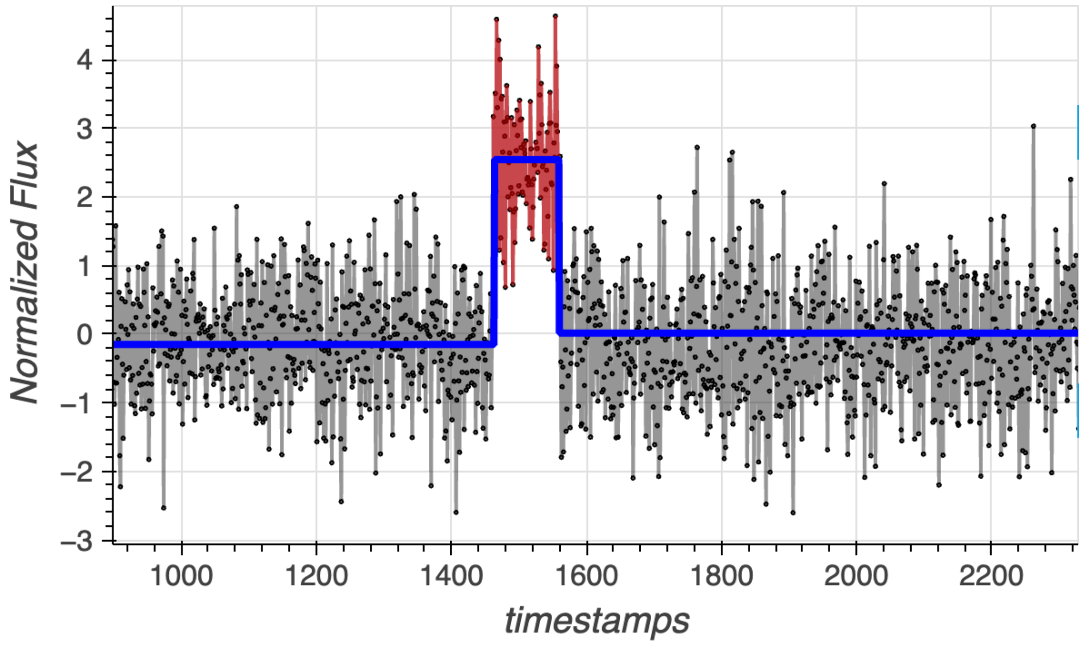

**Unstable behavior**



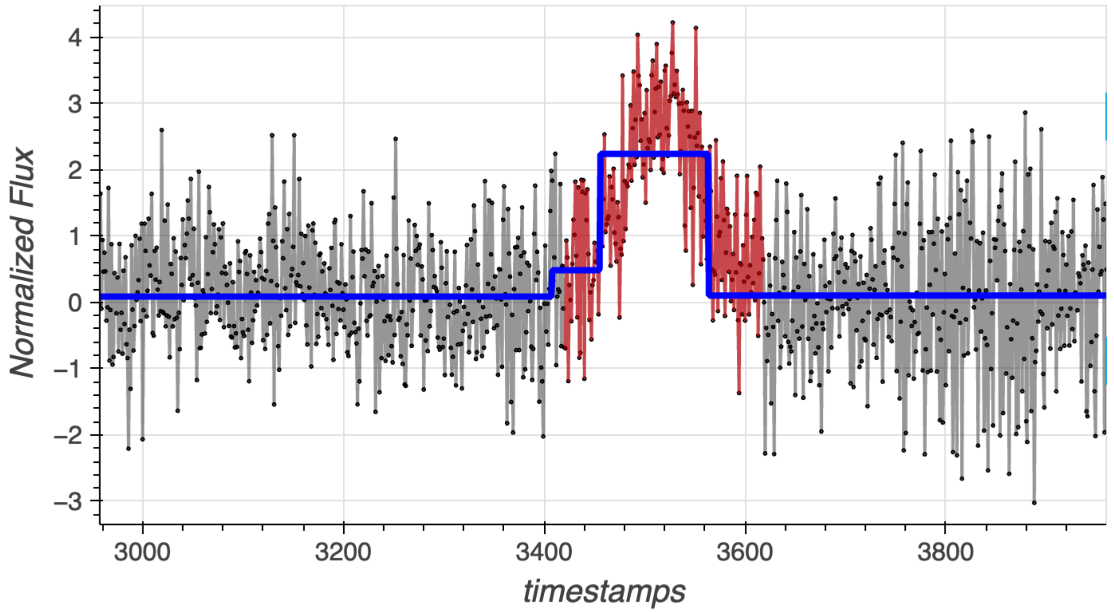

**Gradual change**

# Sketching results using EBinning

EBinning can provide sketching for the many characteristic.



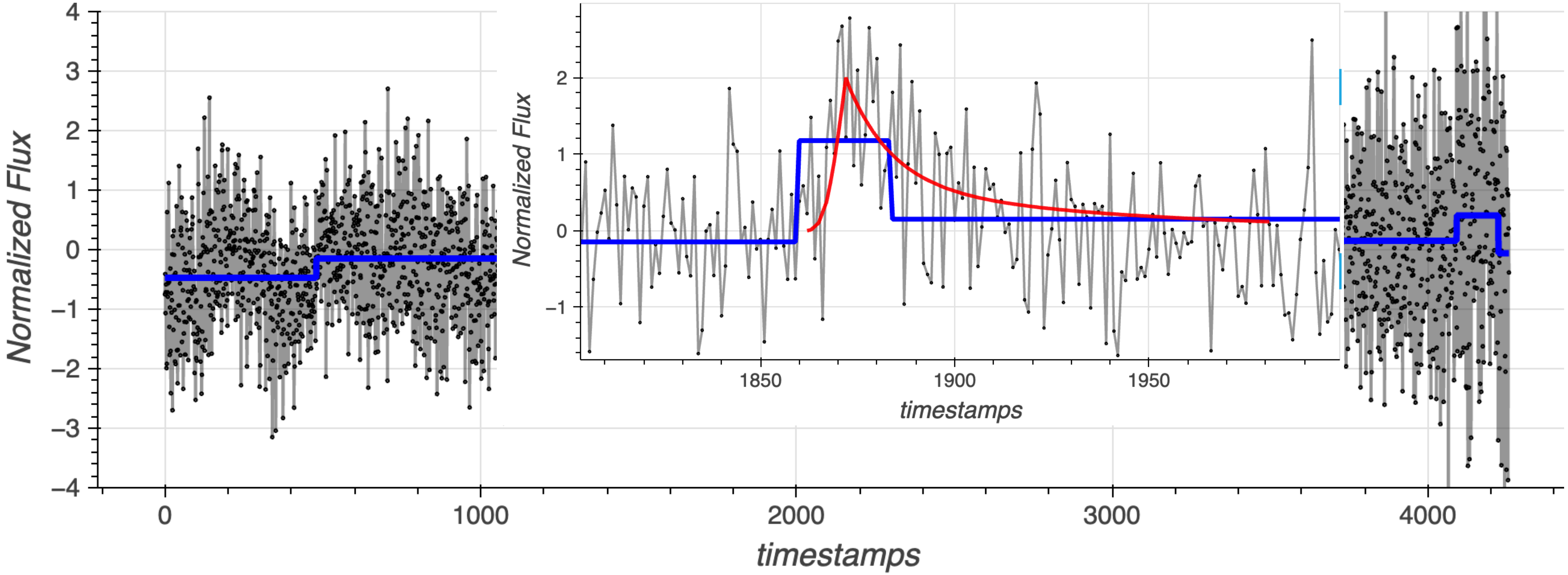

**Square shape**




**Triangle shape**

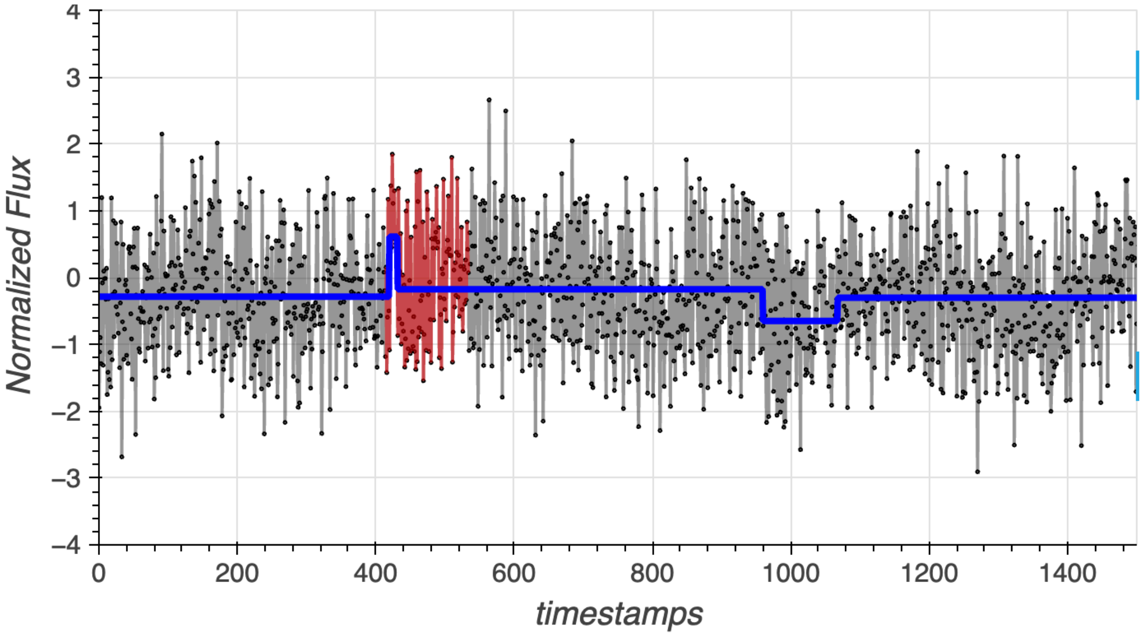
# EBinning for transient pattern detection

When we inspect sketching by eyes, we can identify the sudden change period in a few seconds.

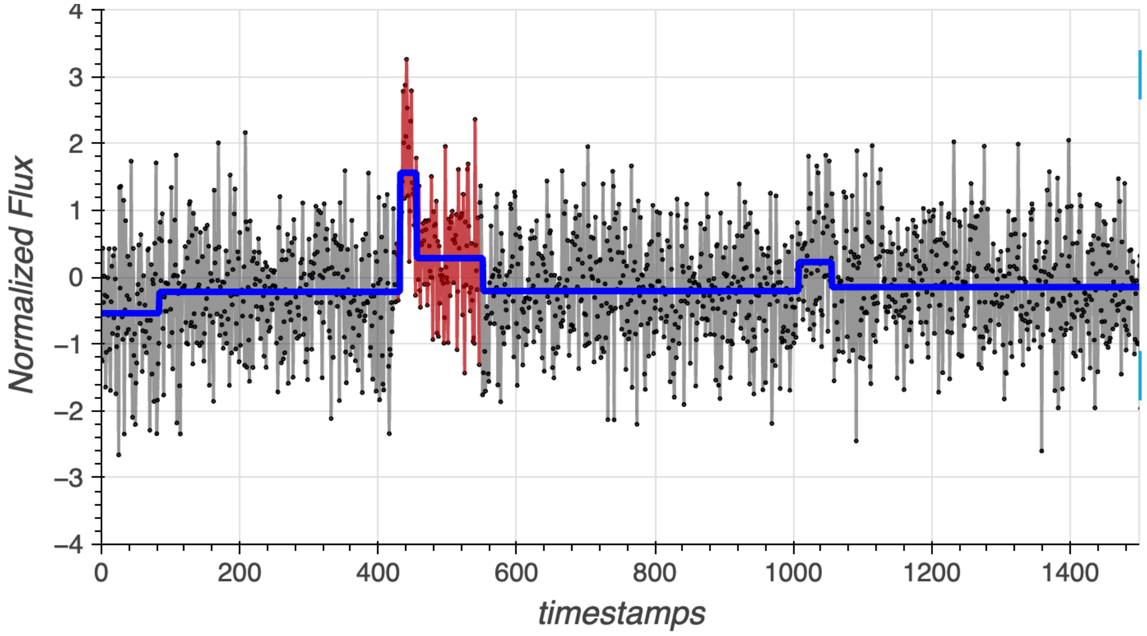


# EBinning for visualization

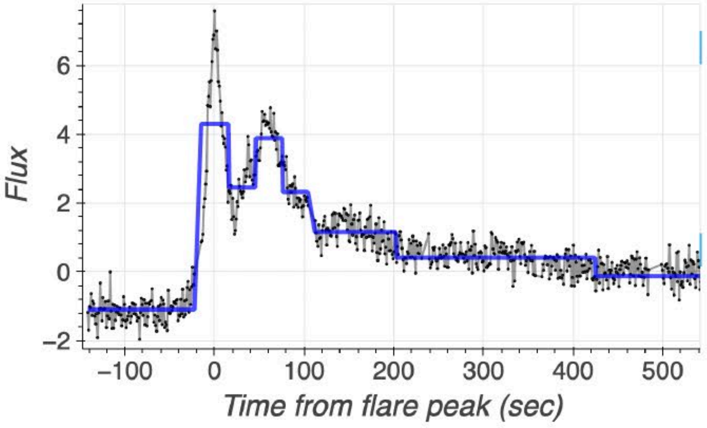
EBinning can provide right sketching for low- and high-power flares.



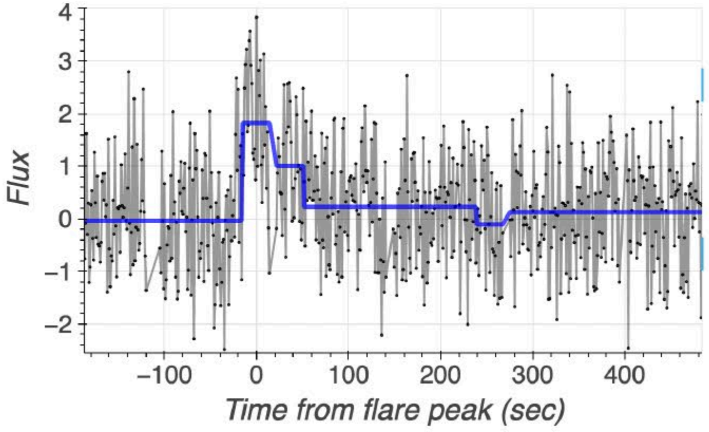

**Low power flare (1sigma)**



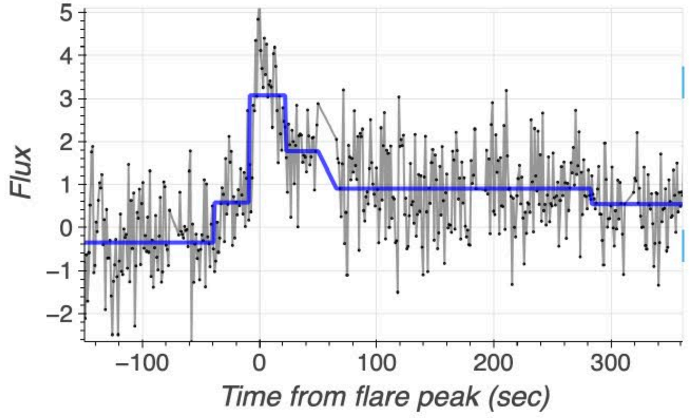

**High power flare (3sigma)**



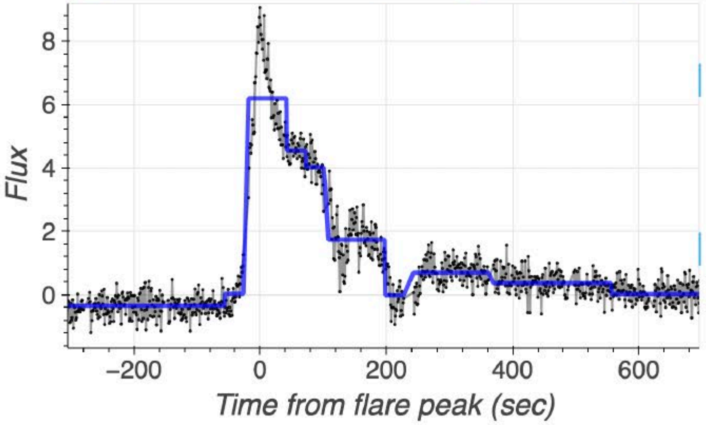

(a) TIC15904458 (Complex flare)



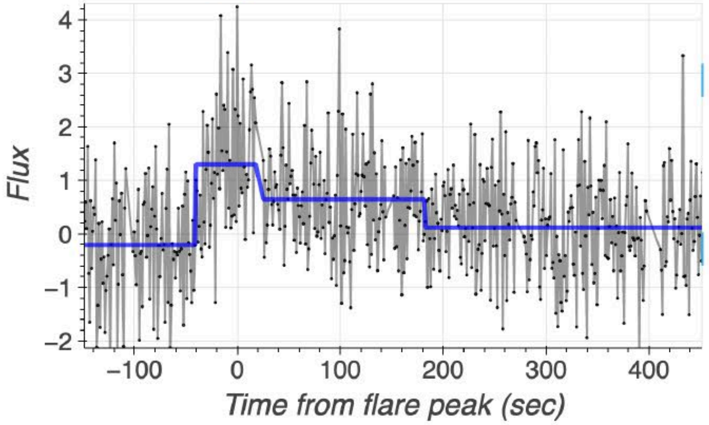

(b) TIC17198188



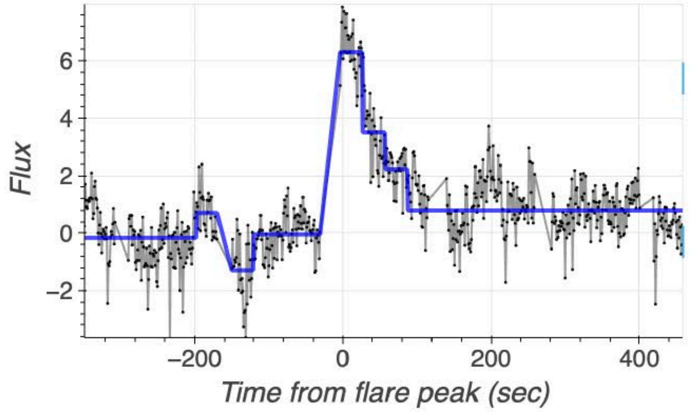

(c) TIC18376490




(d) TIC20536892




(e) TIC21286088

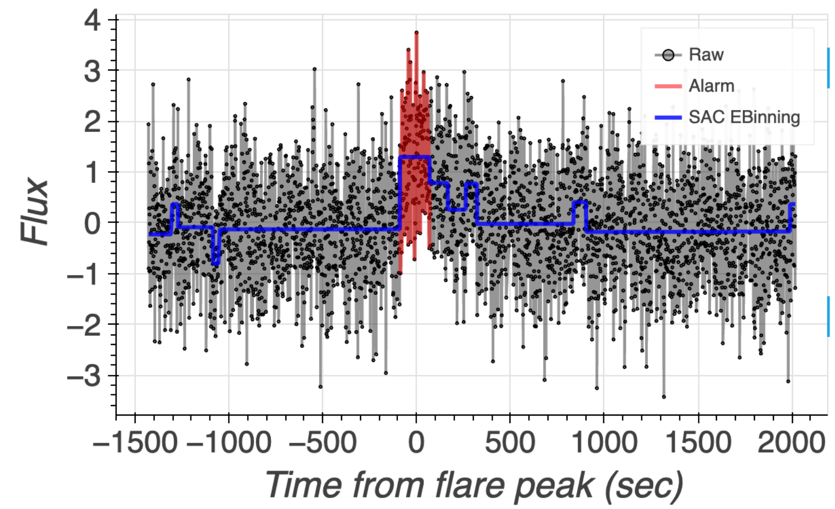



(f) TIC21286574

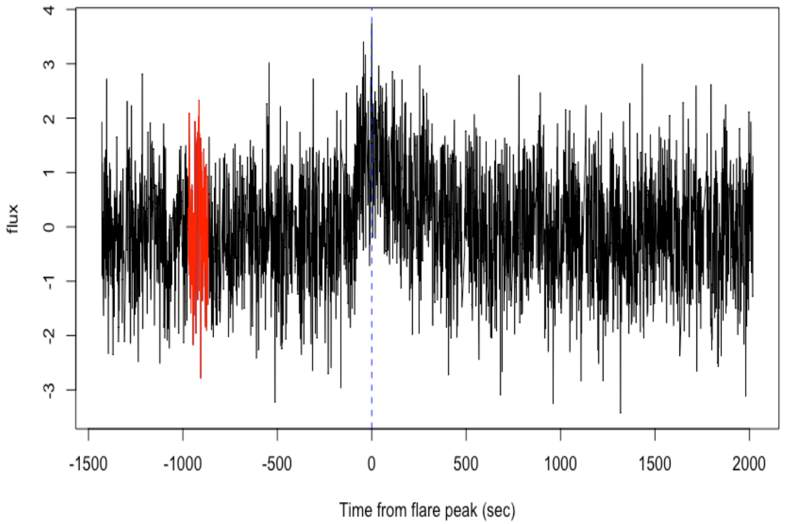
Real flares representation results using EBinning

# Comparison with other methods

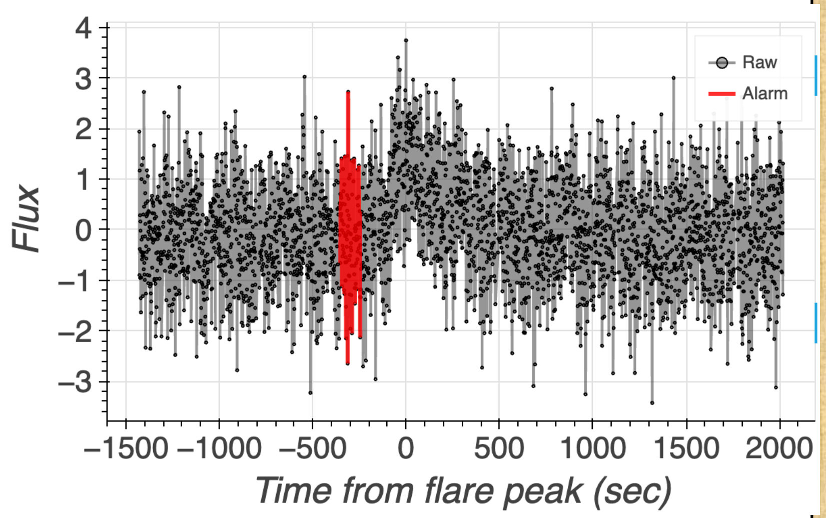
- Our sketch was able to identify the period of the flare correctly.




EBinning




Heuristically Ordered Time Series using Symbolic Aggregate Approximation




Matrix Profile

# Further plan

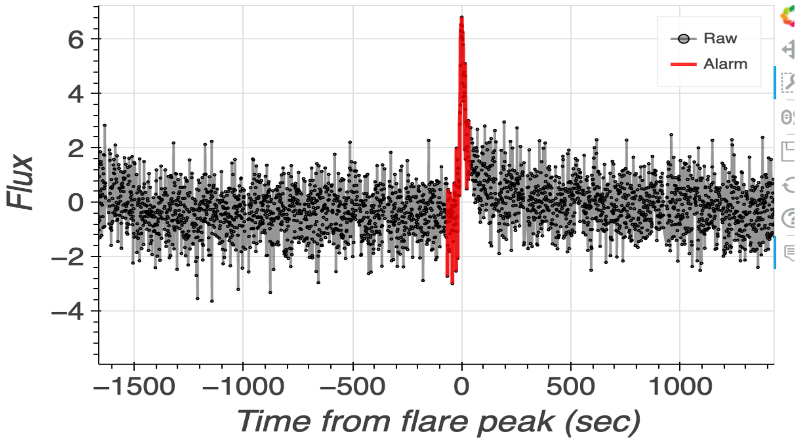
- Creating a package for reproducibility using EBinning in Python and R.
- Creating scripts for reproducibility using alternative methods to analyze lightcurves.
- Applying Large Language Model for analyzing with lightcurves.

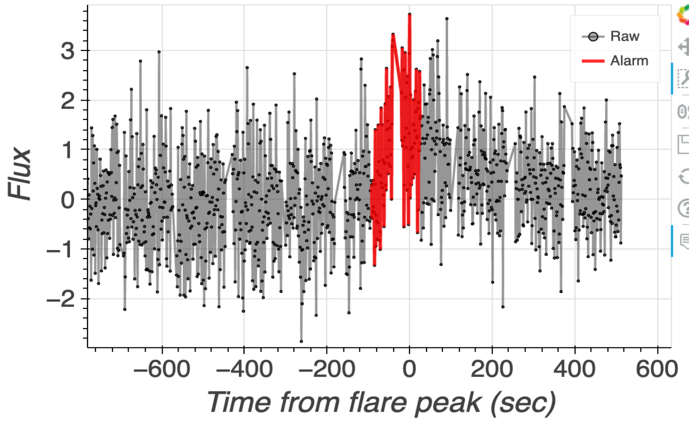
Source code (Python)/
Examples

URL:
https://sites.google.com/view/elasticdatabinning/

# Appendix : Matrix Profile (MP)

- MP algorithm aims to identify motifs and discords in time series data.
  - Motifs refer to repeating patterns or subsequences that appear frequently in the time series.
  - Discords represent unusual or anomalous patterns in the time series.

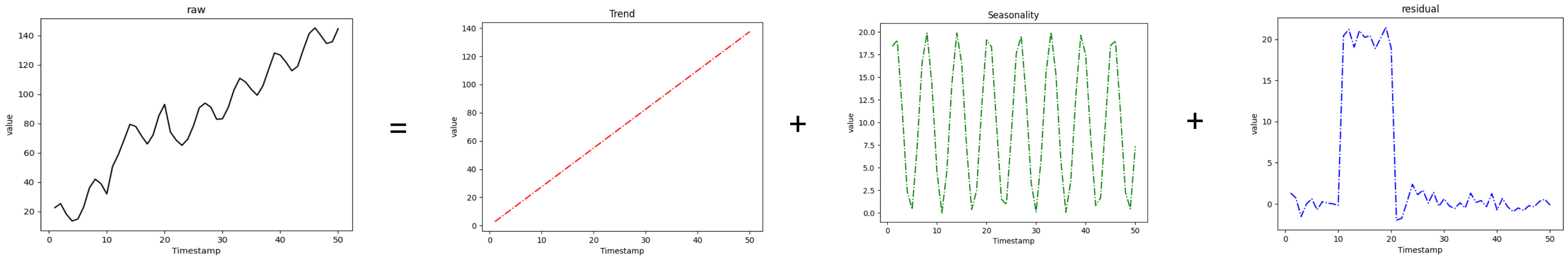



C. -C. M. Yeh et al., "Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets," 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 2016, pp. 1317-1322, doi: 10.1109/ICDM.2016.0179.

# Appendix : Seasonal and Trend Decomposition using Loess (STL)

- STL is a method that separates time series data into seasonal patterns, long-term trends, and residual components.
- The decomposition provided by STL helps remove noise and provides a clearer interpretation of the time series.

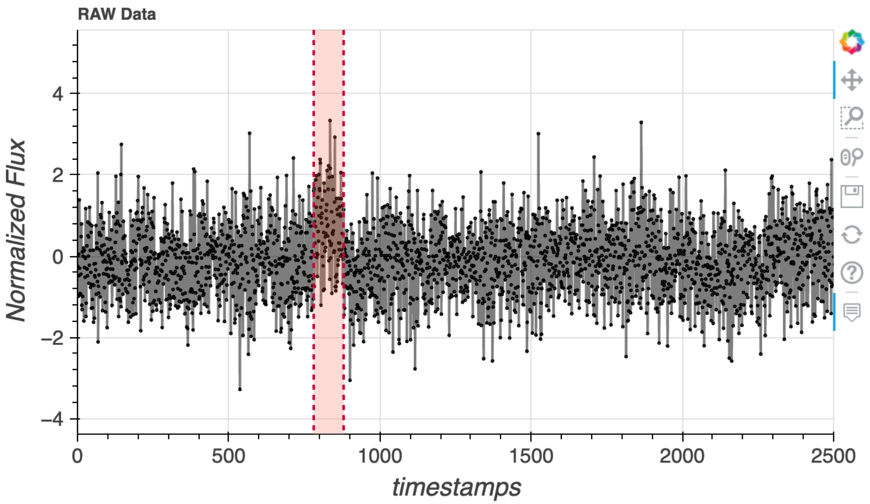



Time series data | Trend | Seasonal | Residual

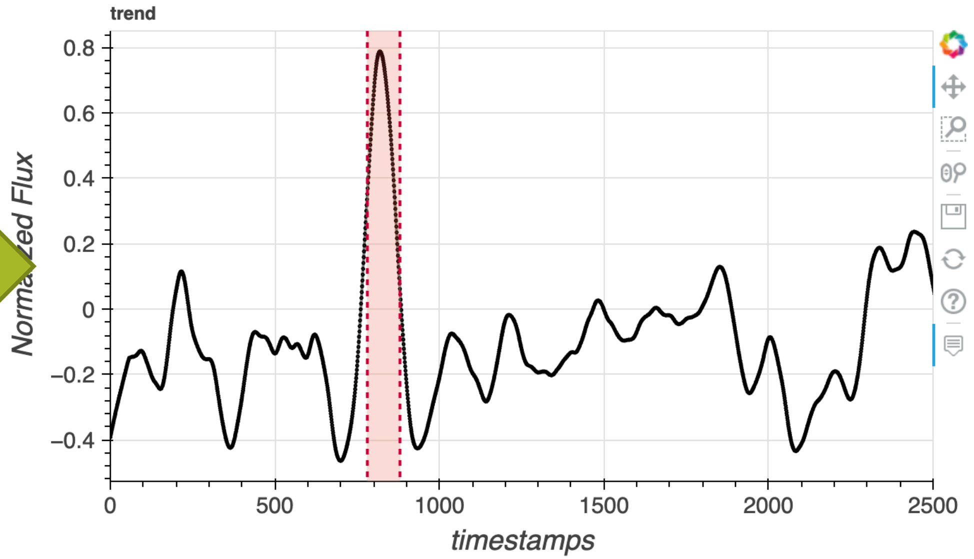
Cleveland, Robert B., et al. "STL: A seasonal-trend decomposition." *J. Off. Stat* 6.1 (1990): 3-73.

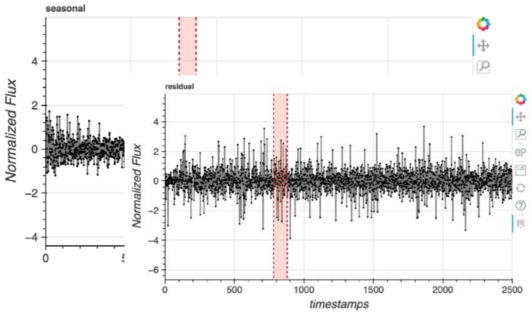
# Appendix : result using STL

Result after decomposition (Trend component)




STL




unwanted components removing

# ありがとうございました。

# Source code (Python)/ Examples

URL:

https://sites.google.com/view/elasticdatabinning/